

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2004, 07:39:56 ; Search time 4765.54 Seconds
(without alignments)

8565.996 Million cell updates/sec

Title: US-09-743-823-21

Perfect score: 1367

Sequence: 1 taaataatctatacatataa.....ctacacacactagaagaata 1367

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estfun:*
16: em_eston:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gse1:*
29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.8	6.1	1200	13	EX437758
2	80	5.9	1200	13	EX415878
3	79.8	5.8	887	13	EX441520
4	79.4	5.8	1201	13	EX446296

RESULT 1

EX437758

LOCUS

DEFINITION

EX437758 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YB01

5-PRIME, mRNA sequence.

ACCESSION

EX437758

VERSION

EST.

KEYWORDS

Source

ORGANISM

Homo sapiens (human)

REFERENCE

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID: CS0CAP008CA01QPI.

Location/Qualifiers

1. .1200

/organism="Homo sapiens"

FEATURES

source

5	79	5.8	1165	13	EX338369
6	77.6	5.7	987	29	CNS014PQ
7	76.6	5.6	1201	13	EX334085
8	76.4	5.6	887	13	EX441520
9	75.4	5.5	1101	29	CNS00K35
10	75.2	5.5	1101	29	CNS0039C
11	75.2	5.5	1201	13	EX446296
12	75.2	5.5	1391	29	CG754863
13	75	5.5	905	29	CNS00K35
14	74.4	5.4	1101	29	CNS00EVL
15	74.4	5.4	1200	13	EX415878
16	74.2	5.4	1201	9	AL536104
17	73.8	5.4	1201	9	AL536104
18	73.6	5.4	1201	9	AL536104
19	73.4	5.4	888	28	EX439779
20	73.4	5.4	928	29	EX439779
21	73.4	5.4	1201	13	EX421282
22	73.2	5.4	681	13	EX378802
23	73	5.3	928	29	CNS00DKY
24	73	5.3	1200	13	EX437758
25	72.8	5.3	1098	13	EX377526
26	72.6	5.3	911	13	EX446782
27	72.6	5.3	1162	9	AL513719
28	72.6	5.3	1201	13	EX463356
29	72.4	5.3	1092	29	CNS020K7
30	72.4	5.3	1201	9	AL561317
31	72.4	5.3	1201	13	EX430226
32	72.2	5.3	1201	13	EX443774
33	72	5.3	886	9	AL514691
34	72	5.3	1201	9	AL532464
35	71.8	5.3	983	13	EX456702
36	71.8	5.3	1201	13	EX406178
37	71.4	5.2	633	13	EX403519
38	71.4	5.2	1101	29	CNS00Z2U
39	71.4	5.2	1201	13	EX360615
40	71.2	5.2	1101	29	CNS00EQL
41	71.2	5.2	1101	29	CNS00EVL
42	70.8	5.2	1098	13	EX377526
43	70.8	5.2	1104	9	AL513809
44	70.6	5.2	1056	13	EX415058
45	70.6	5.2	1201	9	AL532464

ALIGNMENTS

EX437758 1200 bp mRNA linear EST 15-MAY-2003
EX437758 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008YB01
5-PRIME, mRNA sequence.

ACCESSION

EX437758

VERSION

EST.

KEYWORDS

Source

ORGANISM

Homo sapiens (human)

REFERENCE

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. Contact: Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID: CS0CAP008CA01QPI.

Location/Qualifiers

1. .1200

/organism="Homo sapiens"

```

/mc1_type="mrna"
/db_xref="taxon:9606"
/clone="CS0CAP008Y104"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match      6.1%; Score 83.8; DB 13; Length 1200;
Best Local Similarity 31.3%; Pred. No. 0.00037;
Matches 296; Conservative 173; Mismatches 469; Indels 7; Gaps 2;

QY 2 AAAATAATCTATACATATAAAATTTGATTTTAAATTTTAGAAATTCATGATTTTATTT 61
DB 125 AAAAGAAGAAGAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 184
QY 62 TTTTACCAGAAATCCGTTAATATGTTAAATATTACCACTAATTTATATATTTAT 121
DB 185 TTTTATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 244
QY 122 TTTAGGCAATTAAGCATGTTTGATAAATAATATATATTTGTTATTAATCTTTCAAAG 181
DB 245 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 304
QY 182 TATAAGTTGATGAGCGGTGGTGTAGATTTATTTAGTTCTAGTTCTGAAAGCAAGTTG 241
DB 305 ATTGGGGGWTTTTMMWTAAERDXTTTTTTTTAAATAAATAAATAAATAAATAAATA 364
QY 242 GTTTAGACATTTAGCCATTTCTTTTCTACCAAAATAAATGTAATGTAATGTAATGTA 301
DB 365 TTTTATTTTAAWWDDDDAKAWADAGATTTTWTWAGRGARKEDGTTIDTAAADWW 424
QY 302 AGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 361
DB 425 TTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 484
QY 362 AGTGACCAACAATATTAATAAATAAGATTTCTCTACAGTAATAGCGATCTTTTAT 421
DB 485 TTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 544
QY 422 TCAATATCTTTTCCATCTTAAATAATCTTGAGATTTGCACCGTTAACTAAATTAAGTGA 481
DB 545 T-----TTTATTTAAWWATATATTTTWTWAAATAAATAAATAAATAAATAAATA 598
QY 482 TATCCACCGTCTTAAATAAATACTGTGTACCGTGTCTACATTTCAACTTTTGGCACCC 541
DB 599 TTTTACAGGATTTTWTWAAAGDKWTTTWTWTTTWTTTTWTTTTWTTTTWTTTTWT 658
QY 542 GAAAGCCGTATGTTTAGTGTGTCACACAGTTGCAAGCGCATCTACAGAGGCTA 601
DB 659 RGAARETAKATKMTTWTWAAAGCAARKERATWTWAAAWAAGAAAGAGAGA 718
QY 602 CTGGCTCTGCTTTGGCTTTGCTTCTCAATTTTTCATGTTTCTGTTGTTGTTGTTGTT 661
DB 719 AGRAATTTTTTTATKAGARRGATTTTAAWWATATATATATATATATATATATAT 778
QY 662 GTACTTGAACCTATTATATAATATATATATTTTAAAGTTTCTGCTTATATATATCT 721
DB 779 TTTTAAWAAAGATDKAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 838
QY 722 ATATAATATATAGGTTTGAAGATGCAATTTTAAATAAATAAATAAATAAATAAATA 781
DB 839 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 898
QY 782 TAAAAATCGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 841
DB 899 AAAAAAT-AAAAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 957
QY 842 TTTTAAATTTTCCATTTAGATTTCTTTTCTTTAGTTTAAATATGAGCTAGCCAGTTGG 901

```

```

DB 958 WAATAATTTTATTAATATATWAAAAAATTTTWTWAAAAAATTTTATTTATTTATW 1017
QY 902 AGATACACGAAATGTCATGAAACAGTTCATGTCATGCGAAATTTAA 946
DB 1018 TAWTATAGAAATATARDATWRAATTATWADAARAAGAAWATAKDA 1062

RESULT 2
BX415878 1200 bp mRNA linear EST 15-MAY-2003
DEFINITION BX415878 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP008Y104
5-PRIME, mRNA sequence.
ACCESSION BX415878
VERSION BX415878.1 GI:30765550
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/InvitrogenCorporation1600
Faraday Avenue Genoscope sequence ID : CS0CAP008BE02Q1.
Location/Qualifiers.
FEATURES
source
1..1200
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="CS0CAP008Y104"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match      5.9%; Score 80; DB 13; Length 1200;
Best Local Similarity 40.5%; Pred. No. 0.0015;
Matches 134; Conservative 53; Mismatches 144; Indels 0; Gaps 0;

QY 2 AAAATAATCTATACATATAAAATTTGATTTTAAATTTTAGAAATTCATGATTTTATTT 61
DB 854 AATTTTATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 913
QY 62 TTTTACCAGAAATCCGTTAATATGTTAAATATTACCACTAATTTATATATTTAT 121
DB 914 AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAT 973
QY 122 TTTAAGCAATTAAGCATGTTTGCATATAATATATATTTGTTATTAATCTTTTCAAAG 181
DB 974 AATAATATWAAATWAAATWAAATWAAATWAAATWAAATWAAATWAAATWAAATWAA 1033
QY 182 TATAAGTTGATGAGCGGTGGTGTAGATTTATTTAGTTCTAGTTCTGAAAGCAAGTTG 241
DB 1034 AATTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 1093
QY 242 GTTTAGACATTTAGCCATTTCTTTTCTACCAAAATAAATGTAATGTAATGTAATGTA 301
DB 1094 TTTTATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1153
QY 302 AGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 332

```

[illegible]

```

BX338369
LOCUS      1165 bp      mRNA      linear      EST 02-MAY-2003
DEFINITION Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSOD1058YF14 5-PRIME, mRNA sequence.
ACCESSION BX338369
VERSION    BX338369.1 GI:30343682
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 1165)
AUTHORS   Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: segref@genoscope.cns.fr
          Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/InvitroGen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CSOD10580C07QPl.

FEATURES             Location/Qualifiers
     source           1..1165
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CSOD1058YF14"
                     /issue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCWSPORT 6 vector. Library was normalized."

ORIGIN
Query Match      5.8%; Score 79; DB 13; Length 1165;
Best Local Similarity 39.3%; Pred. No. 0.0022;
Matches 130; Conservative 58; Mismatches 143; Indels 0; Gaps 0;

OY 1 TAAATAAATCTATACATTAATAAATTTGATTTTAAATTTTACAAATTCATCTTTATT 60
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 813 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 72
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 61 TTTTTCACGAAATCCGTTAATTTGTTAAATATTTACCAACTAATTTAATAATTTTA 120
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 873 TAAATATATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 932
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 121 TTTTAAGGCAATAGCATGTTGTAATAAATAATATATATTTGTTAATAAATCTTTCAAA 180
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 933 TTTTATTAATAATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 992
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 181 GTATAAGTTGATGCGGTGGTGTAGATTAATTTAGTTCTAGGTTGGAATCAACTT 240
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 993 TATTAATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 1052
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 241 GATTAGCAATTTAGCTTATTTCTTTTCTAACCAATAAATAAATGTAATGAAACCTT 300
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1053 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1112
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 301 TAGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 331
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1113 TTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 1143
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
LOCUS      997 bp      DNA      linear      GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL104456

```

```

VERSION      AL104456.1 GI:5616067
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster (fruit fly)
ORGANISM     Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 987)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
             - web : www.genoscope.cns.fr
COMMENT      Determination of this BAC-end sequence was carried out as part of a
             collaboration with the European Drosophila Genome Project (EDGP) -
             http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
             library (Dros BAC) was made by Alain Billaud at CEPH (Centre
             d'Etude du Polymorphisme Humain) with funding provided by a MRC
             project grant. The DNA was prepared from embryos by Alain Bucheton
             and Genevieve Payan. It has been constructed in the vector
             pBelOBAC11.

FEATURES             Location/Qualifiers
     source           1..987
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN12P22"
                     /clone_lib="DrosBAC"
                     /plasmid="pBelOBAC11"
                     /note="end : SP6"

ORIGIN
Query Match      5.7%; Score 77.6; DB 29; Length 987;
Best Local Similarity 38.3%; Pred. No. 0.004;
Matches 79; Conservative 62; Mismatches 65; Indels 0; Gaps 0;

OY 564 ACTTGAACCTTATTATAAATTACATAATTTTATAAGTTTCCTCTTATATATATCAT 723
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 768 AATTTAAATTTTWTWTATTTTAAATTAATTAATTTATTTATTTATTTATTTAT 827
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 724 ATAATATATAGGTTTAGAATGCCAATTTTAAATAAAGAAATAAATAAATAAATA 783
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 828 ANAKWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 887
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 784 AATTCGAAATAATGAAATGTAATAAATTTAGGGGACAAATAAATAATATGAAAGCTTAT 843
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 888 ATWAAAAAAMWMDRAAAAWDWWWWAAATAAAWTTTAAAAAATAAATAAATAAATA 947
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 844 ATTAAATTTTCCATAGGAATTCAT 869
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 948 WWWWWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 973
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
LOCUS      1201 bp      mRNA      linear      EST 02-MAY-2003
DEFINITION BX334085 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
cDNA clone CSOD06YLO6 5-PRIME, mRNA sequence.
ACCESSION   BX334085
VERSION     BX334085.1 GI:30333283
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE    1 (bases 1 to 1201)
AUTHORS     Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

```

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6562.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DD006DF03QPI&cluster=6562.r>. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0DD006DF03QPI.

FEATURES source

Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DD006Y106"
/tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 5.6%; Score 76.6; DB 13; Length 1201;
Best Local Similarity 45.5%; Pred. No. 0.0051;
Matches 145; Conservative 29; Mismatches 145; Indels 0; Gaps 0;
QY 2 AAAATAATCTATACATTAATAAAATTTGATTTTAAATTTTAGAAATTCATGATTTTATTT 61
Db AAAAAAAAAAAAAAAAAAAATTTATATAAARSGSAAAAAAAAATAATAAAAAAAAAAW 931
QY 62 TTTTTCACGAAATCGTTAATATTTGTTAAATATTACCACTAATTTTAAATTTAT 121
Db WMLIAAAAAAAAAAAAAAAAAATTTTAAAAAAAAAAAAAAAAAAATTTTAAAAAATTTT 991
QY 122 TTTAAGGCAATTAAGCATTTTGTATAAATATATATTTGTATAAATCTTTTCAAAAG 181
Db TTTTAAAAAAAAATTTTTTTAAAAAAAAAAATTTTAAAAAAAAAAAAAAAAAAAA 1051
QY 182 TATAAGTGAATGAGCGTGGTGTAGATATTATTAGTTCTAGTTTGAATGCACTTG 241
Db WAAAAATTTWWTTWW 1111
QY 242 GTTTAGACATTTAGCTTATCTTTTCTAACCAATAATGTAATGTAATGGAACCTTT 301
Db NTT 1171
QY 302 AGGAAAAAAGAAATCAA 320
Db AAAAAAAAAAAATWTATWAA 1190

RESULT 8

EX441520 887 bp mRNA linear EST 15-MAY-2003
LOCUS BX441520 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0DF018YJ11 3-PRIME, mRNA sequence.
ACCESSION BX441520
VERSION BX441520.1 GI:30787976
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 887)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: secref@genoscope.cns.fr, web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7316.r

Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CS0DP018CE06NP1.

FEATURES source

Location/Qualifiers
1. .887
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DP018YJ11"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: Brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 5.6%; Score 76.4; DB 13; Length 887;
Best Local Similarity 42.5%; Pred. No. 0.0066;
Matches 181; Conservative 44; Mismatches 200; Indels 1; Gaps 1;
QY 550 TTATCTTTAGGTAGTTGTTGCAACAGTTGAAGCGCATCTCAGGAGCTACTTGGTCT 609
Db TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 426
QY 610 TCGTTTTCGCTTTTGTTCATTTTTCACGATTTTGTGTCGACACCGCTACTTGA 669
Db KXTT 486
QY 670 ACCTTATTAATTAACATAATTTTAAAGTTTCTTCTTATATAATTAATTAATTAATA 729
Db TT 546
QY 730 TATAGCGTTTAGAATGCCAATTTTATAAAAAAGAAATAAAAAATAAATAAATAAATCG 789
Db TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 606
QY 790 AAAAAATGAATGTAAAAATTTGAGGGGCAAAATAAATAATGAAGTCTATTATTAA 849
Db TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 666
QY 850 ATTTCCTCATTAGATTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 908
Db TTTTWTWTWTAAATTTTWTATTTTAAATAATATATATATATATATATATATATATAT 726
QY 909 CGAAAAATGTCATGAACACAGTTGCTATGTCAGGGAATTAATGTCATGAGGATGAGAAC 968
Db TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 786
QY 969 AAAAAAT 974
Db AAAAAAT 851

RESULT 9

CNS00KBS 1101 bp DNA linear GSS 03-JUN-1999
LOCUS CNS00KBS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BAC39P05 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL077453.1 GI:4956930
VERSION AL077453
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoxer in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers

source 1..1101

organism="Drosophila melanogaster"

molecule="genomic DNA"

db_xref="taxon:7227"

clone_lib="BACRP9805"

clone_lib="RPCL-98"

note="end : T7"

ORIGIN

Query Match 5.5%; Score 75.4; DB 29; Length 1101;

Best Local Similarity 48.2%; Pred. No. 0.0084;

Matches 148; Conservative 10; Mismatches 149; Indels 0; Gaps 0;

QY 666 TTGAACATTATTAATAATACATAATTTTATAGTTTCATCTTTATATATACATAT 725

DB 183 TTTAAATATATAATAATAATAATTAATAATAATAATAATAATAATAATAATA 242

QY 726 ATATATATAGGTTTACATGCGAATTTTAAATAAGATATAATAATAATAATAATA 785

DB 243 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 302

QY 786 ATCGAAAAATGAAATGTAAAAATTTGAGGGGCAATATAATAATGAAATCTATTAT 845

DB 303 ATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 362

QY 846 TTAATATTTCCATAGAAATCTATTTCCTTAGTTAATATAGCTAGCGAGTGGAGAT 905

DB 363 TAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 422

QY 906 ACACGAAATGTCATGAAACAGTTGATGAGGAAATTAATGATGAGGATAGCAA 965

DB 423 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 482

QY 966 GACAAA 972

DB 483 AAAAAA 489

RESULT 10

CNS0039G

LOCUS

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BAC08K10 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL063921

VERSION AL063921.1 GI:4941778

KEYWORDS Drosophila melanogaster (fruit fly)

SOURCE Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

JOURNAL Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoxer in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers

source 1..1101

organism="Drosophila melanogaster"

molecule="genomic DNA"

db_xref="taxon:7227"

clone_lib="BACRP9805"

clone_lib="RPCL-98"

note="end : TET3"

ORIGIN

Query Match 5.5%; Score 75.2; DB 29; Length 1101;

Best Local Similarity 18.5%; Pred. No. 0.009;

Matches 130; Conservative 288; Mismatches 278; Indels 8; Gaps 1;

QY 14 ACATATAAATTTGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 73

DB 396 ATATAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 455

QY 74 AATCGGTTATATTTGTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 433

DB 456 AATATATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 515

QY 134 AAGCATGTTTGAATAAATATATATATTTTAAATTTTAAATTTTAAATTTTAA 193

DB 516 AAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 575

QY 194 GATGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 253

DB 576 TTTTTHYTTTTHYTTTTHYTTTTHYTTTTHYTTTTHYTTTTHYTTTTHYTTTTH 635

QY 254 AGCCTATTTCTTTTCTAACCAATAATAATGTAATGGAACCTTTAGCAAAAAAAG 313

DB 636 TAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 695

QY 314 A-----AATCAAAATTTGAAATCATCTCCGGTGGAGTCGAGAGCCACACCC 365

DB 696 HTCHCYVYHYHTTTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHY 755

QY 366 GACCAACAATATAATAATAAGATTTGCTCTACAGTAATGCGTACTTTTATTCAA 425

DB 756 TWYAAHAAMWMMHHYAAAAAATAATTHYHHYTHYTHYTHYTHYTHYTHYTHYTH 815

QY 426 TACTTTTCCACTTCTAAATCTTTGAGATTTTCCCGTTCATTAATTAAGTGTATATC 485

DB 816 WHYHYHTATCTTTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHY 875

QY 486 CAACGGTCTTAAAAAATCTGTCTACCGCTCACTTTCACTTTGCGACCCCTGAAA 545

DB 876 CWACMTNGHHMMHHHNA CHAHHHTTNCMCHHHCTCHHHHTMYTMTCHNMMH 935

QY 546 GCCTTATGTTAGGTAGTGTGTTTGAACAGTTGAAGCCATCCTCAGGAGCTACTT 605

DB 936 HWHHWAATTTTTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHYTHY 995

[illegible]

Search completed: February 24, 2004, 11:07:12
Job time : 4772.54 secs

Result No.	Query #			DB	ID	Description
	Score	Match	Length			
1	64.4	4.7	6106	14	US-10-311-455-1446	Sequence 1446, App
2	64.4	4.7	6106	16	US-10-257-166-114	Sequence 114, App
3	64.2	4.7	7560	14	US-10-311-455-1196	Sequence 1196, App
C 4	63.4	4.6	7195	14	US-10-239-676-30	Sequence 30, App
	63.4	4.6	7195	14	US-10-240-453-40	Sequence 40, App
C 5	63.4	4.6	7195	14	US-09-814-353-6389	Sequence 6389, App
C 6	63.2	4.6	318	10	US-09-814-353-12666	Sequence 12666, App
C 7	63.2	4.6	318	10	US-09-814-353-17949	Sequence 17949, App
C 8	63.2	4.6	403	10	US-09-814-353-12666	Sequence 12666, App
C 9	63.2	4.6	416	9	US-09-960-352-4584	Sequence 4584, App
C 10	62.2	4.6	375	9	US-09-960-352-15014	Sequence 15014, App
C 11	61.2	4.5	5511	14	US-10-311-455-1974	Sequence 1974, App
C 12	61	4.5	385	10	US-09-814-353-19050	Sequence 19050, App
C 13	60.4	4.4	7631	14	US-10-311-455-833	Sequence 833, App
C 14	60.2	4.4	6487	12	US-10-221-613-227	Sequence 227, App
C 15	60	4.4	6254	14	US-10-311-455-1594	Sequence 1594, App


```
2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 30
; LENGTH: 7195
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-30

Query Match
Best Local Similarity 4.6%; Score 63.4; DB 14; Length 7195;
Matches 157; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

675 ATTATAATTACATAAATTTTATAGTTTCACCTCTTATATATATATATATATATAG 734
6759 ATTATCCAAAAACAATCAATTTTAAATTTTATTTTAAACACTATATATTTTAA 6700
6759 GGTTTAGAAATGCCAATTTTAAAAAAAGAAATAAAAAATAATAGAAATAAAATCCAAAA 794
6699 TATTTTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6640
6759 ATGAAATGTAAAAATTTTGGGGGACAAATAAATAATATGAAGTCTATATTTTAAATTTT 854
6639 AAAACATTTTAAATAATTTTCAAAAAACAAAAAATAATTTTAAAAAATAAATCTTAAAC 6580
6759 CCATTTAGAAATCTTATTTTCTTATTTTCTTATTTATATGAGGTACCGAGTGGGAGATACG 914
6579 CCAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6520
6759 TGTCATGAAACAGTTGCGATGTAGGGGAAATTAATGTAGTAGAGGGATAGCAAGACAAAAAT 974
6519 ACTAACGAATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 6460
6759 CCAAGCCCAAGCTA 987
6459 ACCCAACAAAAA 6447

RESULT 6
US-09-814-353-6389/c
; Sequence 6389, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814.353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6389
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 121, 123, 124, 125, 126, 129, 134, 157, 158, 159, 160, 161,
; LOCATION: 172, 179, 180, 183, 194, 205, 210, 211, 213, 216, 219, 220,
; LOCATION: 225, 226, 227, 235, 241, 243, 248, 250, 267, 268, 272, 277,
; LOCATION: 282, 287, 293, 302
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-6389

Query Match
Best Local Similarity 4.6%; Score 63.2; DB 10; Length 318;
Matches 98; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
```

QY 666 TTGAACTTATATATACATTAATTTTATAGTTTCACCTCTTATATATACATCAT 725
Db 200 TTATATATTTTAAATAAANNNTTTTAAATAATTTNNNNNTTTTAAATAAATAA 141
QY 726 AATATATAGGTTAGAAATGCCAATTTTAAATAAGATAAATAAATAATAGATAA 785
Db 140 AAAAAATGGGNCNNNTNTTTTAAATAAATAAATAAATAAATAAATAAATAA 81
QY 786 ATCGAAAAATGAATGTAAATAATTCAGGGGCAATATAAATAATGAAGT 838
Db 80 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAGT 28

RESULT 7

US-09-814-353-12666/c
; Sequence 12666, Application US/09814353
; Publication No. US2003016583A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR FILING DATE: 2000-06-15
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12666
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 121, 123, 124, 125, 126, 129, 134, 157, 158, 159, 160, 161,
; LOCATION: 172, 179, 180, 183, 194, 205, 210, 211, 213, 216, 219, 220,
; LOCATION: 225, 226, 227, 235, 241, 243, 248, 250, 267, 268, 272, 277,
; LOCATION: 282, 287, 293, 302
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-12666

Query Match 4.6%; Score 63.2; DB 10; Length 318;
Best Local Similarity 56.6%; Pred. No. 0.08;
Matches 98; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 666 TTGAACTTATATATACATTAATTTTATAGTTTCACCTCTTATATATACATCAT 725
Db 200 TTATATATTTTAAATAAANNNTTTTAAATAATTTNNNNNTTTTAAATAAATAA 141
QY 726 AATATATAGGTTTGAATGCCAATTTTAAATAAGATAAATAAATAATAGATAA 785
Db 140 AAAAAATGGGNCNNNTNTTTTAAATAAATAAATAAATAAATAAATAAATAA 81
QY 786 ATCGAAAAATGAATGTAAATAATTCAGGGGCAATATAAATAATGAAGT 838
Db 80 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAGT 28

RESULT 8

US-09-814-353-17949
; Sequence 17949, Application US/09814353
; Publication No. US2003016583A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR FILING DATE: 2000-06-15
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17949
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 225
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-17949

Query Match 4.6%; Score 63.2; DB 10; Length 403;
Best Local Similarity 53.3%; Pred. No. 0.087;
Matches 131; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 603 TTGGTCTTGCTTTGGCTCTTTTGTCAATTTTTCACGTGATTTTGTGTCGACACGCG 662
Db 115 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 174
QY 663 TACTGAACTTATATAAATACATAATTTATAGTTTCACCTCTTATATATACATCA 722
Db 175 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 234
QY 723 TATAATATAGGTTTAGAATGCCAATTTTAAATAAGATAAATAAATAATAGATA 782
Db 235 AAAAAAATTTTATGGGATAAAAAATTTTAAATAAATAAATAAATAAATAAATAA 294
QY 783 AATATCGAAAAATGAATGTAAATAATTCAGGGGCAATATAAATAATGAAGTCTAT 842
Db 295 AAAAACTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 354
QY 843 TATTT 847
Db 355 AATTT 359

RESULT 9

US-09-960-352-4584
; Sequence 4584, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Ningbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4584
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 20-LTB3057-016-Q1-K1-E11
US-09-960-352-4584

Query Match 4.6%; Score 62.2; DB 9; Length 416;
Best Local Similarity 51.8%; Pred. No. 0.088;
Matches 169; Conservative 0; Mismatches 153; Indels 4; Gaps 1;
QY 603 TTGCTGCTTTTGGCTTTTGGTCAATTTTTCAGTCAATTTTGTGGTGAACGCG 662
Db 5 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 64
QY 663 TACTTGAACTTATTATAATACATAATTTTATAAGTTTCACCTTCTTATATACTCA 722
Db 65 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 124
QY 723 TATAATATATAGGTTTAGAATGCAATTTTAAAGAAAGAAATATAAGAAATATAAGAT 782
Db 125 AATGAATTTTAAAGAAAGAAATATAAGAAAGAAATATAAGAAAGAAATATAAGAA 184
QY 783 AATATCGAAAGAAATGTAAGAAATTTGAGGGGACAAATATAAGAAATATAAGAA 838
Db 185 AGATTTTAAAGAAATATAAGAAATATAAGAAATATAAGAAATATAAGAAATATAAGAA 244
QY 839 CTATTATTAAATTTCCATAGAAATCTATTTCCTTTAGTTTATATAGCTAGCCAGTT 898
Db 245 AAAAAATTTTAAATTTTAAAGAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAT 304
QY 899 GGGATACACGAAATGTCATGAAA 924
Db 305 AAGATATAAGAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAG 330

RESULT 10
US-09-960-352-15014/c
; Sequence 15014, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 15014
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 64-LTB3058-048-Q1-K1-H8
US-09-960-352-15014

Query Match 4.6%; Score 62.2; DB 9; Length 375;
Best Local Similarity 53.0%; Pred. No. 0.13;
Matches 133; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 619 GTCTTTGCTCAATTTTTCAGTCAATTTTGGTGAACGCGTACTTGAACCTTATTA 678
Db 276 GTATTTTAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 217
QY 679 TAAATACATAATTTTATAAGTTTCACCTTCTTATATAATCTCATATAATATATAGGGTT 738
Db 216 TATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 157

QY 739 TAGAATGCCAATTTTAAAAAAGAAATAAAAAAATAAATAGAAATCGAAAAATGA 798
Db 156 AATAAAAAAATAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 97
QY 799 AATGTAAAGAAATTTGAGGGGACAAATTAATATGAAAGTCTATTATTAAATTTTCCAT 858
Db 96 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 37
QY 859 TAGAATTTCTAT 869
Db 36 AAAAAAATTTAT 26

RESULT 11
US-10-311-455-1974/c
; Sequence 1974, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCI/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1974
; LENGTH: 5511
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1974

Query Match 4.5%; Score 61.2; DB 14; Length 5511;
Best Local Similarity 51.9%; Pred. No. 0.54;
Matches 163; Conservative 0; Mismatches 148; Indels 3; Gaps 1;
QY 674 TATTATAATACATAATTTTATAGTTTCACTTCTTATATATATATATATATATATA 733
Db 2933 TAAAAATTTTAAAAAATCAATACATATTCACATATTTTATATATTTTCAATATATA 2874
QY 734 GGGTTTGAATGCCAATTTTAAAAAAGAAATAAAAAATAAATAGAAATCGAAAA 793
Db 2873 AATAATAATACATAATTTTAAACAATAAAAACATTTAAATAAATAAATCAATTTCTAT 2814
QY 794 AATGAAATGTAAGAAATTTGAGGGGACAAATATAAGAAATGAAAGTCTATTATTAAATTT 853
Db 2813 AACGAAATAAAAAACATCTTAAA---AAAAATAAAAAACACAAATATAATTTTAACTA 2757
QY 854 TCCATTAGAAATTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 913
Db 2756 TTAATAAATAATCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2697
QY 914 ATGTCATGAACAGTTGCAATGAGGAAATTAATGATGAGGGATAGCAAGACAAAA 973
Db 2696 ATTCTTTTAAATATATTTTAAAAAATAAATTTTATACATAAATAATCAATAAATA 2637
QY 974 TCCAGCCCAAGCTA 987
Db 2636 ATATACCTTATTTA 2623

RESULT 12
US-09-814-353-19050/c

QY 249 CATTAGCCTTATCTCTTTTCTAACCAATAATGAAACCTTTAGGAAA 308
Db 4321 GTTTTTTAAATAAAGGCGTTTCTTTTAAATAATGCTGGAGTGATACGG 4380
QY 309 AAAAGAAATCAAAATTTGAAACATCATCCGGTGA 343
Db 4381 AATAGTATTAGTAGAGCAATATAATAATAGGA 4415

RESULT 15

```

US-10-311-455-1594
Sequence 1594, Application US/10311.455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PISPENBROOK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
TITLE OF INVENTION: cytosine methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1594
LENGTH: 6254
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1594

```

Query Match	4.4%;	Score 60;	DB 14;	Length 6254;
Best Local Similarity	53.9%;	Pred. No. 0.93;		
Matches 166;	Conservative 0;	Mismatches 140;	Indels 2;	Gaps 2;

QY	9	TCTATACATATAAAAAATTCGATTTTAAAAATTTAGAAATTCATGATTTTATTTTTTTT	68
Db	2641	TTTTTATATAAGAGATTGGTTTTTAATAATTTTAAATTAAGAAATTTTTTTTAAAGTA	2700
QY	69	CCGAAATCCGTTAATATTTGTTAAAAATATTACCAACTAATTTATAAAATTTTATTTAAAGG	128
Db	2701	TAGCGATAATTTAAATTTTITTAATTTAGTAATTAAGACTATAAAATTTTATTTAGAATT	2760
QY	129	CAATTAGCATGTTGTATAAAATATATATTTGTATATAATCTTTTCAAAAGATAAAG	188
Db	2761	AAAGTAATAATTTTTTG-TAATTATAAATATAATTTAAATTTTAATAAATATATTATTAAT	2819
QY	189	TTGATGATGGCGTGGTAGATTAATTTA-CTTCTAGGTTCCGAATCGCACTTGGTTTAG	247
Db	2820	TTGAGTCGAGTATTTATTTGTATAATAATATATTTTTTTTTTTTGTATGAGAATGTTTTA	2879
QY	248	ACATTTAGCCCTATTCCTTTTTCTTAACCCAAAATAAAATGTAATCGAAAACCTTTAGGAAA	307
Db	2880	TTAGGTAAATAATGATTAATTAAGAACTTTTAAAGTAAATGATAGCAGGAGATGTCAGAT	2939
QY	308	AAAAAGAA	315
Db	2940	ATAAGAA	2947

Search completed: February 24, 2004, 13:09:15
Job time : 614.765 secs

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2004, 07:40:35 ; Search time 134.614 Seconds
(without alignments)
5635.487 Million cell updates/sec

Title: US-09-743-823-21

Perfect score: 1367

Sequence: 1 taaataatctatcatataa.....ctacacacactagaagaata 1367

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 692709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
C 1	59	4.3	6124	4	US-08-213-419B-3		Sequence 3, Appl1
C 2	56.4	4.1	5332	4	US-09-801-861-3		Sequence 3, Appl1
C 3	55	4.0	2246	4	US-09-363-708-3		Sequence 3, Appl1
C 4	55	4.0	2246	4	US-09-083-587-3		Sequence 3, Appl1
C 5	54.8	4.0	640681	4	US-09-780-988-1		Sequence 1, Appl1
C 6	54.2	4.0	5852	1	US-07-857-106-2		Sequence 2, Appl1
C 7	53.6	3.9	615	3	US-08-998-416-186		Sequence 186, App
C 8	53.6	3.9	636	3	US-08-998-416-1137		Sequence 1137, App
C 9	53.6	3.9	837	3	US-08-998-416-288		Sequence 288, App
C 10	53.4	3.9	8537	4	US-10-204-708-41		Sequence 41, Appl
C 11	52.6	3.8	6124	4	US-08-213-419B-3		Sequence 3, Appl1
C 12	52.2	3.8	827	3	US-08-998-416-535		Sequence 535, App
C 13	52.2	3.8	20674	4	US-09-641-638-851		Sequence 631, App
C 14	51.6	3.8	6306	4	US-10-204-708-49		Sequence 49, Appl
C 15	51.6	3.8	8920	2	US-08-446-855A-1		Sequence 1, Appl1
C 16	51.6	3.8	8920	3	US-08-150-741-1		Sequence 1, Appl1
C 17	51	3.7	662	3	US-08-998-416-185		Sequence 185, App
C 18	51	3.7	663	3	US-08-998-416-191		Sequence 191, App
C 19	51	3.7	665	3	US-08-998-416-937		Sequence 937, App
C 20	51	3.7	701	3	US-08-998-416-701		Sequence 701, App
C 21	51	3.7	711	3	US-08-998-416-786		Sequence 786, App
C 22	51	3.7	724	3	US-08-998-416-683		Sequence 683, App
C 23	51	3.7	732	3	US-08-998-416-1036		Sequence 1036, App
C 24	51	3.7	767	3	US-08-998-416-472		Sequence 472, App
C 25	51	3.7	828	3	US-08-998-416-538		Sequence 538, App
C 26	51	3.7	834	3	US-08-998-416-305		Sequence 305, App
C 27	50.8	3.7	6317	4	US-10-204-708-11		Sequence 11, Appl1

28	50.4	3.7	6243	2	US-09-056-075-1	Sequence 1, Appl1
C 29	50.4	3.7	8537	4	US-10-204-708-42	Sequence 42, Appl1
C 30	50.4	3.7	19124	2	US-08-487-826B-13	Sequence 13, Appl1
C 31	50.4	3.7	26000	4	US-09-843-376-10	Sequence 10, Appl1
C 32	50.2	3.7	8093	4	US-10-204-708-31	Sequence 31, Appl1
C 33	50	3.7	8669	4	US-10-204-708-6	Sequence 6, Appl1
C 34	50	3.7	7218	1	US-08-232-463-14	Sequence 14, Appl1
C 35	49.8	3.6	640681	4	US-09-790-988-1	Sequence 1, Appl1
C 36	49	3.6	194	4	US-09-621-976-9596	Sequence 9596, App
C 37	49	3.6	4673	1	US-07-638-431-1	Sequence 1, Appl1
C 38	49	3.6	4673	5	PCT-US92-00018-1	Sequence 1, Appl1
C 39	49	3.6	6866	4	US-10-204-708-20	Sequence 20, Appl1
C 40	49	3.6	7304	4	US-10-204-708-44	Sequence 44, Appl1
C 41	49	3.6	19233	4	US-10-204-708-46	Sequence 46, Appl1
C 42	48.8	3.6	1117	3	US-09-247-373B-33	Sequence 33, Appl1
C 43	48.8	3.6	1361	4	US-09-489-847-64	Sequence 64, Appl1
C 44	48.8	3.6	6070	4	US-10-204-708-9	Sequence 9, Appl1
C 45	48.8	3.6	11015	4	US-10-204-708-55	Sequence 55, Appl1

ALIGNMENTS

RESULT 1
US-08-213-419B-3/c
Sequence 3: Application US/08213419B
Patent No. 6333406
GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: J11-002CNCIP
; CURRENT APPLICATION NUMBER: US/08/213,419B
; PRIOR FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match 4.3%; Score 59; DB 4; Length 6124;
Best Local Similarity 56.4%; Pred. No. 0.0043;
Matches 110; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 675 ATTATAAATACATATTTTATAGTTTCACCTCTTATATAATATCTCATATATATATAG 734
DB 782 ATATTAAATATAAAGAAATACATATATAGTTTAACTAAGACTAGTTTATTCATAT 723
QY 735 GGTTTAGAAATGCAATTTTATAAAGAAATATAAATAAATAGATATAAATCGAAAAA 794
DB 722 TTATAATATTAAATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 663
QY 795 ATGAATGTAAAAAATTTGAGGGGACAAATATAAATGAAAGTCTTATTATTATTTT 854
DB 662 ATAAATAGAAACAAACTGTATATAAATTTTCATATGATTATTCATTATTATTA 603
QY 855 CCATTAGAAATTCCTAT 869
DB 602 TCATAAAATTTTAT 588

DESCRIPTION: /desc = "mouse PAL cDNA"
US-09-083-587-3
Query Match 4.0%; Score 55; DB 4; Length 2246;
Best Local Similarity 57.1%; Pred. No. 0.022;
Matches 100; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 662 GTACTGAACTTATTATTAATACATAATTTTATAAGTTTCACTTCTTATATAATACATC 721
Db 2049 GTATTGACATGCTTAAAGTATGCTTATATATATCTTTCATCTTGCCTCATGGCT 2108
QY 722 ATATATATATAGGTTAGAAATGCCAATTTTAAAAAAGATATAAATAATAGAA 781
Db 2109 GTGTATTATATAAGTGTACTTGACCAAAAAAATAAATAAATAAATAAATAA 2168
QY 782 TAAATCGAAAAATGAATGTAAAAATTTGAGGGGCAATAAATAATATGAAA 836
Db 2169 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2223
RESULT 5
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790, 988
; PRIOR FILING DATE: 2001-02-23
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1
Query Match 4.0%; Score 54.8; DB 4; Length 640681;
Best Local Similarity 42.9%; Pred. No. 0.061;
Matches 381; Conservative 0; Mismatches 502; Indels 5; Gaps 2;
QY 1 TAAAAATACTATACATAAAAAATTTGATTTTAAATTTAGAAATTCATGATTTATT 60
Db 353537 TTAAGTAATATCAATTTTAACTAATATTTAAAAATTTATATGTTTCTAAAAATAGATA 353478
QY 61 TTTTATACAGAAATCCGTTATATTTTAAATATTTACCACTA---ATTATTAATT 117
Db 353477 AATTTTATTAATCTTTTATTTTATTTTAAATTAACATAGGTATGATTAAGA 353418
QY 118 TTATTTTAAAGCAATTAAGCATGTTTGATATAATATATATTTGTTATTAATACTTTTCA 177
Db 353417 AATCAATTAAGTAATGAATGATTTATTTTAAATTAACATATAAATAAATAAATAA 353358
QY 178 AAGTATAAAGTTGATCATGGCGGTGGTGGATTTTATTTAGTCTAGGTTCGAATGCAA 237
Db 353357 GAAATATTTTCGTATGATCATATATTTTAAAGGTATTTCTATTAATAAATAATATGATA 353296
QY 238 GTTGGTTTACATTTAGCCCTTATCTTTTCTTACCAAAAAATAATGTAATGGAAC 297
Db 353297 AGAATTTAAATAAACTTTTATTTATTTTATTTTATTTATTTTAAATTTATCTT 353238
QY 298 CTTTAGAAAAAAGAAATCAAAATTTGAAACATCATCGGTGGAGTCGAGAGCCAC 357
Db 353237 TTCTGGTATTATTCTTAATAGATAAATAAATAAATAAATAAATAAATAAATAAATAA 353178
QY 358 ACCACGTGACCCCAACAATATTAATAAATAAGAGTTTGTCTACAGTAAATCCGATATCTTT 417

Db 353177 AAAAAATGATAATCTCTTTTATAAGATTAATCTTTTATTTTATTTTATTTATTTATTTT 353118
QY 418 TTATTCATATCTTTTCCACTTCATTAATCTTTGGAGATTTTGCACCGTTAACTAATTAAGT 477
Db 353117 TTATTCATATTAAGTATCTTTTAAAAAATATGATTAATAATAATTTTAAAAATTTAT 353058
QY 478 GTTATATCCAAAGGTCCTRAAAAAACTTGTGTACGGTCCCTCACATTTCAACTTTGGCCA 537
Db 353057 TATTAATATCTTCCACAAAAATTAATAAGAAATATCTTTTAAACATAAATCAATCT 352998
QY 538 CCGTCAAGCCGTTATGTTTAGGTTAGTGTTCGCAACAGTTTGAAGCGCATCACTCAGGAG 597
Db 352997 TAATTTGAAGTAAGTATCACTTATAGTCAATATTAACAGTTTAACTATTTACTACTAT 352938
QY 598 GCTACTTGGCTTTCGCTTTTGGCTTTTGTTCATTTTTCACGTTGATTTTGGTGAAC 657
Db 352937 ATTTTAAAAAATGTTTATTTTAAATTTCTGAAATTTAGTATTTATTTTATTTTATGAA 352878
QY 658 ACGGTACTTGAACCTTATTAATAATTACATAATTTTATATAAGTTTCACTTCTTATATAAT 717
Db 352877 TAGGCATATTTTATATATAAAGTATGTTTAAATACTTTTGTAGTATAGATT 352818
QY 718 ACTCATATAATATAGGTTTGAAGTCCAAATTTTAAAAAAGAAATAAAAAATAAAT 777
Db 352817 TTTATAATATAATTAAGTCTATATAAATTTCTGTTAATTTGATTAATTTTATAACA--TATC 352760
QY 778 AGATAAAAAATCGAAAAAATGAATGTAAAAAATTTGAGGGGCAATAAATAATATGAAG 837
Db 352759 CCAGTATTTTCAATATTTTGTGTAACTAACGCTGGAACAAATAAATAATTTGAAAA 352700
QY 838 TCTATTATTTAAATTTTCCATTAGAAATCTTATTTTCTTCTAGTTAATAT 885
Db 352699 TCAATACATTTGTTTTCATATAATATAATTTTATTTTATTTTATTTTATTTTAAAT 352652

RESULT 6
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

Query Match
Best Local Similarity 4.0%; Score 54.2; DB 1; Length 5852;
Matches 112; Conservative 0; Mismatches 78; Indels 1; Gaps 1;

QY 683 TTACATAATTTTAAAGTTTCACCTTCCTTATA-TAATACCTCATATAAATATAGGGTTTAG 741
DB 5713 TTACTTATATTTCTATTTTATTTTATTAATAATATATTAATTTTAAATAAATAAAAA 5654

QY 742 AATGCCAATTTTAAAAAAGAAATAAAAAATAATAGATAAAATCGAAAAATGAAAT 801
DB 5653 AAAAAAATAAAAAAATAATTTAAATTTAAAAAATAAAAAAATAAAAAAATAAAAA 5594

QY 802 GTAAAAAATTTGAGGGGACAAATAAATATCGAAAGTCTATTATTAAATTTTCCATAG 861
DB 5593 ATGTATAATTTGGAATTAATAAATAAATGTAAGGGGTTTTTTTAAATAATGATGTCG 5534

QY 862 AATCTATTTT 872
DB 5533 ATTTTATTTT 5523

```

```

RESULT 7
US-08-998-416-186
; Sequence 186, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jorgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE: 31-DEC-1996
; APPLICATION NUMBER: CH 0016/97
; NAME: Meigs, J. Timothy
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 38,241

```

```

; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: FAG1074RP
; US-08-998-416-186

Query Match
Best Local Similarity 3.3%; Score 53.6; DB 3; Length 615;
Matches 107; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

QY 5 ATAATCTATACATTAAAAAATTTGATTTTAAAAATTTTAGAAATTCATGATTTTATTTT 64
DB 26 ATAAGATTATATAAATCTTTTATATATAATTTTAAAGTATTAATTTTAAACTATT 85

QY 65 TTTACGAGAAATCCGTTAAT--ATTGTTAAATATTACCACTAAATTTTAAATTTTAT 122
DB 86 ATTATCATTTTAAATAAATAATTAATTTTGAATTAATTAATTTTATATAATTTATA 145

QY 123 TTAAGGCAATTAAGCAATGTTGATAAAATATATATATTTGTTATAAATACATTTTCAAA 180
DB 146 TAATTACTTAATTCATCATTTATATATTTATATTAATTTATAAATAATTTATA 203

RESULT 8
US-08-998-416-1137
; Sequence 1137, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jorgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689

```

```
; INFORMATION FOR SEQ ID NO: 1137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1692RP
US-08-998-416-1137

Query Match          3.9%; Score 53.6; DB 3; Length 636;
Best Local Similarity 60.1%; Pred. No. 0.035;
Matches 107; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

QY 5 ATAATCTATACATTAATAAATTTGATTTTAAAAATTTTAGAAATTCAGATTTTATTTT 64
    |||||
Db 26 ATAAAGATTAAATAAATTTTATTAATAATTTTAAAGTATTAAATTTTAAACTATT 85
    |||||

QY 65 TTACACAGAAATCCGTTAAT--ATTGTTAAATAATACCACTAAATTTATAAATTTTATT 122
    |||||
Db 86 ATATCATTTTAAATAAATTTTATTAATAATTTTGAATTTATTAATACTATTATAATTATA 145
    |||||

QY 123 TTAAGGCAATTAAGCATGTTTGATAAATAATATATATTTGTTATAAATACTTTTCAAAA 180
    |||||
Db 146 TAAATTACTTAATCAATTAATTAATTTTATTAATAATTAATAAATAAATAAATA 203
    |||||

RESULT 9
US-08-998-416-288
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSVPII
; NUMBER OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PP/5-30306/A/CGCI976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
US-08-998-416-288

Query Match          3.9%; Score 53.6; DB 3; Length 837;
Best Local Similarity 60.1%; Pred. No. 0.036;
Matches 107; Conservative 0; Mismatches 69; Indels 2; Gaps 1;

QY 5 ATAATCTATACATTAATAAATTTGATTTTAAAAATTTTAGAAATTCAGATTTTATTTT 64
    |||||
Db 26 ATAAAGATTAAATAAATTTTATTAATAATTTTAAAGTATTAAATTTTAAACTATT 85
    |||||

QY 65 TTACACAGAAATCCGTTAAT--ATTGTTAAATAATACCACTAAATTTATAAATTTTATT 122
    |||||
Db 86 ATATCATTTTAAATAAATTTTATTAATAATTTTGAATTTATTAATACTATTATAATTATA 145
    |||||

QY 123 TTAAGGCAATTAAGCATGTTTGATAAATAATATATATTTGTTATAAATACTTTTCAAAA 180
    |||||
Db 146 TAAATTACTTAATCAATTAATTAATTTTATTAATAATTAATAAATAAATAAATA 203
    |||||

RESULT 10
US-10-204-708-41/c
; Sequence 41, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 41
; LENGTH: 8537
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-204-708-41

Query Match          3.9%; Score 53.4; DB 4; Length 8537;
Best Local Similarity 53.6%; Pred. No. 0.058;
Matches 111; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 667 TGAACCTTATTAATAAATTTACATAATTTTATAGTTTCTCTTATATAAATCTCATATA 726
    |||||
Db 962 TAAAAATATTTATAAATATATTAATTAATAAATTTTAAAAATATATAAATCTCACACTTATA 903
    |||||

QY 727 ATATATAGGTTTACAGATTCGCAATTTTAAAAAAGATAAAAAATTAATAGATAATAA 786
    |||||
Db 902 AICCTAATACATTAATAAATCTAAATAAATAAATAAACCATTAAACCAAAATTAATAACTA 843
    |||||

QY 787 TCGAAAAAATGAATGCTAAAAATTTGAGGGGACAAATAAATAATGAAGTCTATTATT 846
    |||||
Db 842 ATCTAAACAAATATAAATAAATAAATTAATTAATTAATTAATTAATTAATTAATTTT 783
    |||||

QY 847 TAAATTTCCATTAGAAATTCATTTC 873
```



```

; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/205,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 651
; LENGTH: 20674
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1123..3123
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 3124..3297
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 3871..4072
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 5552..5633
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 5758..5880
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 5996..6099
; OTHER INFORMATION: exon 5
; NAME/KEY: exon
; LOCATION: 6349..6509
; OTHER INFORMATION: exon 6
; NAME/KEY: exon
; LOCATION: 7379..7522
; OTHER INFORMATION: exon 7
; NAME/KEY: exon
; LOCATION: 8645..8854
; OTHER INFORMATION: exon 8
; NAME/KEY: exon
; LOCATION: 12254..12340
; OTHER INFORMATION: exon 9
; NAME/KEY: exon
; LOCATION: 12854..13023
; OTHER INFORMATION: exon 10
; NAME/KEY: exon
; LOCATION: 13308..13429
; OTHER INFORMATION: exon 11
; NAME/KEY: exon
; LOCATION: 16567..16667
; OTHER INFORMATION: exon 12
; NAME/KEY: exon
; LOCATION: 16775..16945
; OTHER INFORMATION: exon 13
; NAME/KEY: exon
; LOCATION: 17063..17554
; OTHER INFORMATION: exon 14
; NAME/KEY: misc_feature
; LOCATION: 17555..20674
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 1128
; OTHER INFORMATION: 10-508-191 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1182
; OTHER INFORMATION: 10-508-245 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1559
; OTHER INFORMATION: 10-509-284 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1570
; OTHER INFORMATION: 10-509-295 : deletion of C
; NAME/KEY: allele
; LOCATION: 1827
; OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTTT
; NAME/KEY: allele
; LOCATION: 2048
; OTHER INFORMATION: 10-511-62 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2323
; OTHER INFORMATION: 10-511-337 : insertion of T
; NAME/KEY: allele
; LOCATION: 2341
; OTHER INFORMATION: 10-512-36 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 2623
; OTHER INFORMATION: 10-512-318 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2832
; OTHER INFORMATION: 10-513-250 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2844
; OTHER INFORMATION: 10-513-262 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2934
; OTHER INFORMATION: 10-513-352 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2947
; OTHER INFORMATION: 10-513-365 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 3802
; OTHER INFORMATION: 12-206-81 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 4062
; OTHER INFORMATION: 10-343-231 : deletion of C
; NAME/KEY: allele
; LOCATION: 4088
; OTHER INFORMATION: 12-206-366 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4109
; OTHER INFORMATION: 10-343-278 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4170
; OTHER INFORMATION: 10-343-339 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 5903
; OTHER INFORMATION: 10-346-23 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6019
; OTHER INFORMATION: 10-346-141 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6141
; OTHER INFORMATION: 10-346-263 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6183
; OTHER INFORMATION: 10-346-305 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6338
; OTHER INFORMATION: 10-347-74 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6375
; OTHER INFORMATION: 10-347-111 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6429
; OTHER INFORMATION: 10-347-165 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6467
; OTHER INFORMATION: 10-347-203 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6484
; OTHER INFORMATION: 10-347-220 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6534
```



```
/
/
/  ZIP: 22201-4714
/  COMPUTER READABLE FORM:
/  MEDIUM TYPE: Floppy disk
/  COMPUTER: IBM PC compatible
/  OPERATING SYSTEM: PC-DOS/MS-DOS
/  SOFTWARE: Patent Release #1.24
/  CURRENT APPLICATION DATA:
/  APPLICATION NUMBER: US/08/446,855A
/  FILING DATE: 06-Jul-1995
/  CLASSIFICATION: 435
/  ATTORNEY/AGENT INFORMATION:
/  NAME: Mitchard, Leonard C
/  REGISTRATION NUMBER: 29,009
/  REFERENCE/DOCKET NUMBER: 47-80
/  TELECOMMUNICATION INFORMATION:
/  TELEPHONE: 703-816-4000
/  TELEFAX: 703-816-4100
/  INFORMATION FOR SEQ ID NO: 1:
/  SEQUENCE CHARACTERISTICS:
/  LENGTH: 8920 base pairs
/  TYPE: nucleic acid
/  STRANDEDNESS: single
/  TOPOLOGY: linear
/  MOLECULE TYPE: genomic
/  US-08-446-855A-1

Query Match      3.8%; Score 51.6; DB 2; Length 8920;
Best Local Similarity 53.5%; Pred. No. 0.13;
Matches 108; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 662 GTACTTGAACTTATTATTAATACATAATTTTATTAAGTTTCACCTTCTTATATATACCTC 721
Db 526 GTTCCCTAATAATTTATTATATATATTTATTATTAATTTATTTATTTATTTTTC 585

QY 722 ATATAATATATAGGTTTAGAATCCCAATTTTAAAAAAGATAAAAAATAAATAGAA 781
Db 586 TTAGTTTATAAATAGTAATTCCTACTAATTTAAAAAATAAATAAATAAATAAATAA 645

QY 782 TAAATCGAAATGAATGTAATAAATTTGAGGGGACAAATAAATATGAAGTCTA 841
Db 646 GAAAAAATAAATAAATTTACATATGAAAAATGAACTTGATATATGTAATTTATATAATTT 705

QY 842 TTATTTAAATTTTCCATTAGAA 863
Db 706 AACAATAAATAAATGATAAA 727
```

Search completed: February 24, 2004, 11:09:40
Job time : 138.614 secs

BEST AVAILABLE COPY

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2004, 06:53:45 ; Search time 705.304 Seconds
(without alignments)
8233.739 Million cell updates/sec

Title: US-09-743-823-21

Perfect score: 1367

Sequence: 1 taaataatctatcattataa.....ctacacacacacagaagaata 1367

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_297an04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002s.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1367	100.0	1367	3	AZ58125	Soybean 7
2	1282.8	93.8	1428	9	ACF80579	Basic 7S
3	74.6	5.5	8056	7	ABZ10246	Haematopo
4	72.6	5.3	8056	7	ABZ10246	Haematopo
5	71.4	5.2	8056	7	ABZ10100	Haematopo
6	64.4	4.7	6106	6	AAS46430	Tumour su
7	64.4	4.7	6106	6	ABK40032	Human che
8	64.4	4.7	6106	6	ABL33473	Human imm
9	64.2	4.7	7560	6	ABL33473	Human imm
10	64	4.7	309	5	ABV44994	Human pro
11	63.4	4.6	7195	6	AAS45325	Chemical
12	63.4	4.6	7195	6	ABK28166	DNA trans
13	63.2	4.6	416	7	ABX39419	Bovine ES
14	63.2	4.6	8056	7	ABZ10100	Haematopo
15	62.2	4.6	375	7	ABX49849	Bovine ES
16	62	4.5	115218	7	ACB48485	Human HNR
17	61.2	4.5	626	5	ABV60941	Human pro
18	61.2	4.5	5511	6	ABL34001	Human imm
19	60.4	4.4	7631	6	ABL32860	Human imm
20	60.2	4.4	6487	6	ABK31362	Signal tr
21	60.2	4.4	6487	6	ABV70319	Human gen
22	60.2	4.4	6487	6	AAS61266	Chemical
23	60	4.4	6254	6	ABL33621	Human imm

C	24	59.8	4.4	416	7	ABX39419	Bovine ES
	25	59.8	4.4	1501	7	ABZ10188	Haematopo
	26	59.8	4.4	1501	9	ABE84162	Human lym
	27	59.6	4.4	15743	6	ABK28396	DNA trans
	28	59.2	4.3	626	5	ABV60941	Human pro
	29	59.2	4.3	7319	6	ABL34045	Human imm
C	30	59	4.3	6124	2	AAQ03568	Sequence
	31	59	4.3	6124	6	ABK15042	Plasmodiu
C	32	59	4.3	110000	6	ABX06316_14	Continuation (15 o
	33	58.8	4.3	430	4	AAI89813	Human pol
	34	58.8	4.3	5499	6	ABQ66972	Human ang
	35	58.8	4.3	7145	6	ABL92234	Chemical
	36	58.8	4.3	7145	6	AAD22321	Chemical
	37	58.8	4.3	11155	6	ABL32605	Human HNR
	38	58.8	4.3	115218	7	ACA64845	Human HNR
C	39	58.4	4.3	2000	7	ADA71938	Rice gene
	40	58.4	4.3	13326	6	ABL33713	Human imm
	41	58.2	4.3	6912	6	ABK28372	DNA trans
	42	58	4.2	425	4	AAI85301	Human pol
	43	58	4.2	4233	7	ABZ10136	Haematopo
	44	58	4.2	4233	7	ABZ10240	Haematopo
	45	58	4.2	4233	9	ABE84150	Human lym

ALIGNMENTS

RESULT 1
AAZ58125...
ID AAZ58125 standard; DNA; 1367 BP.
XX
AC AAZ58125;
XX
DT 25-APR-2000 (first entry)
XX
DE Soybean 7S basic globulin gene promoter.
XX
KW Globulin; promoter; soybean; lactoferrin; human; transgenic plant;
KW seed storage protein; ds.
XX
OS Glycine max.
XX
PN W0200004146-A1.
XX
PD 27-JAN-2000.
XX
PF 19-JUL-1999; 99WO-IT000226.
XX
PR 17-JUL-1998; 98IT-RN000478.
XX
(PLAN-) PLANTECHNO SRL.
XX
FI Fogher C;
XX
DR WPI; 2000-161129/14.
XX
PT Synthetic polynucleotide encoding human lactoferrin, used for production
of functional foods, vegetal milks and human lactoferrin.
XX
FS Claim 7; Page 85-86; 93pp; English.
XX
The present sequence is that of the promoter region of the soybean var.
"High and seed-specific 7S basic globulin gene". The promoter can be
incorporated into an expression cassette of the invention also including
a synthetic gene (see AAZ58122) encoding human lactoferrin and designed
for expression of human lactoferrin in transgenic plants. Use of
particular leader sequences and promoters allows expression of the
lactoferrin in specific tissues, especially seeds. The transgenic plants
can be used in processes for the production of functional vegetal milk,
fruit juices, fruit and/or vegetable homogenized foods (claimed).
XX
Expression of lactoferrin in plants should overcome problems associated
with microbial expression and simplify purification. Expression in seeds
also leads to an increase in iron content of seed products. The

CC transgenic plants are selected from solanaceae, cereals, leguminosae,
 CC fruit bearing plants and horticultural plants, especially soybean,
 CC tobacco and rice
 XX
 SQ Sequence 1367 BP; 475 A; 245 C; 202 G; 445 T; 0 U; 0 Other;

Query Match 100.0%; Score 1367; DB 3; Length 1367;
 Best Local Similarity 100.0%; Pred. No. 6.4e-212;
 Matches 1367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATAATCTATACATATAAAATTTGATTTTAAATTTTAGAAATTCATGATTTTAT 60
 DB 1 TAAATAATCTATACATATAAAATTTGATTTTAAATTTTAGAAATTCATGATTTTAT 60

QY 61 TTTTATACAGAAATCCGTTAATATTGTTAAATATTACCACTAATTTATATAATTTTA 120
 DB 61 TTTTATACAGAAATCCGTTAATATTGTTAAATATTACCACTAATTTATATAATTTTA 120

QY 121 TTTTAAGCAATTAAGCAATTTGATAAATAATATATATTGTTATAAATACATTTTCAAAA 180
 DB 121 TTTTAAGCAATTAAGCAATTTGATAAATAATATATATTGTTATAAATACATTTTCAAAA 180

QY 181 GTATAAGTTGATGAGCGGTGGGTAGATTATTTTGTCTAGGTTGCAATGCAAGTT 240
 DB 181 GTATAAGTTGATGAGCGGTGGGTAGATTATTTTGTCTAGGTTGCAATGCAAGTT 240

QY 241 GGTTTAGACATTTAGCCCTTATTTCTTCTAACCAAAATAATGTAATGGAACCTT 300
 DB 241 GGTTTAGACATTTAGCCCTTATTTCTTCTAACCAAAATAATGTAATGGAACCTT 300

QY 301 TAGAAAAAAGAAATCAAAATGAAAACATCATCCGGTGGACTCGAGAAGCCACAC 360
 DB 301 TAGAAAAAAGAAATCAAAATGAAAACATCATCCGGTGGACTCGAGAAGCCACAC 360

QY 361 CACGTGACCAACAAATATAAATAAGAGTTTGTCTACAGTAATGCGATATCTTTTAA 420
 DB 361 CACGTGACCAACAAATATAAATAAGAGTTTGTCTACAGTAATGCGATATCTTTTAA 420

QY 421 TTCAATATCTTTTCCACTTTTAAATCTCGAGATTGCAACCGTTAACTAATTAAGTGT 480
 DB 421 TTCAATATCTTTTCCACTTTTAAATCTCGAGATTGCAACCGTTAACTAATTAAGTGT 480

QY 481 ATATCCACCGCTTAAATAAATCTGTACCGTGCCTACATTTCAATTTTGGCAGCC 540
 DB 481 ATATCCACCGCTTAAATAAATCTGTACCGTGCCTACATTTCAATTTTGGCAGCC 540

QY 541 TGAAGCGGTTATGTTAGTGTAGTTGTCACAGTTGAAGCGCATCACTCAGGAGCT 600
 DB 541 TGAAGCGGTTATGTTAGTGTAGTTGTCACAGTTGAAGCGCATCACTCAGGAGCT 600

QY 601 ACTTGGTCTTCTTTTGGCTTTTGTTCATTTTTCAGTGATTTTGTGGTGAACAG 660
 DB 601 ACTTGGTCTTCTTTTGGCTTTTGTTCATTTTTCAGTGATTTTGTGGTGAACAG 660

QY 661 CGTACTTGAACCTTATATATAAATACATAATTTTAAAGTTTCACITCTTATATAACT 720
 DB 661 CGTACTTGAACCTTATATATAAATACATAATTTTAAAGTTTCACITCTTATATAACT 720

QY 721 GATATAATATAGGTTTGAATGCCAATTTTAAATAAAGATAAATAAATAATAGTA 780
 DB 721 GATATAATATAGGTTTGAATGCCAATTTTAAATAAAGATAAATAAATAATAGTA 780

QY 781 ATAAATCGAAAAATCAATGTATAAATAATTTGAGGGGCAATAAATAATCAAGTCT 840
 DB 781 ATAAATCGAAAAATCAATGTATAAATAATTTGAGGGGCAATAAATAATCAAGTCT 840

QY 841 ATTAATTAATTTTCCATTAGATTTCTATTTTCTTGTAGTTAATAGCTAGCCAGTTG 900
 DB 841 ATTAATTAATTTTCCATTAGATTTCTATTTTCTTGTAGTTAATAGCTAGCCAGTTG 900

QY 901 GAGATACAGAAATGTATGCAACACGTTGCATGTAGGAAATTAATGTAGTAGAGGAT 960
 DB 901 GAGATACAGAAATGTATGCAACACGTTGCATGTAGGAAATTAATGTAGTAGAGGAT 960

RESULT 2

ACF80579 standard; DNA; 1428 BP.

AC ACF80579;

XX 15-JAN-2004 (first entry)

DT

XX Basic 7S soy globulin promoter.

DE

XX Soybean; plant; globulin; promoter; lysosome; enzyme; transgenic plant;
 KW ds.

XX Glycine max.

XX WO2003073839-A2.

XX 12-SEP-2003.

XX 03-MAR-2003; 2003WO-IT000120.

XX 01-MAR-2002; 2002IT-RM000115.

XX (PLAN-) PLANTECHNO SRL.

XX Fogher C, Reggi S;

XX WPI; 2003-712829/67.

XX

XX New genetically transformed plant that can produce a lysosomal enzyme of
 PT animal or human origin, useful for preparing a medicament for enzyme
 PT replacement therapy in Gaucher, Anderson-Fabry or Pompe disease.
 XX Disclosure; Page 47; 53pp; English.

CC The present sequence is that of the basic 7S soy globulin promoter.
 CC Expression cassettes including this promoter may further comprise a DNA
 CC sequence coding for a lysosomal enzyme and a signal sequence capable of
 CC dispatching the lysosomal enzyme to the seed storage organs of a plant.
 CC The lysosomal enzyme is preferably human glucocerebrosidase, human alpha-
 CC galactosidase or human alpha-glucosidase. The invention is based on the
 CC discovery that lysosomal enzymes can be expressed in seed storage organs
 CC in a form which is stable (over 12 months in stored seeds), enzymatically

CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferation disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match 5.5%; Score 74.6; DB 7; Length 8056;
 Best Local Similarity 44.6%; Pred. No. 0.0029;
 Matches 393; Conservative 0; Mismatches 474; Indels 14; Gaps 2;

QY 9 TCTATACATTAATAAATTCGATTTAAATTTAGAAATTCATGATTTATTTTATTTTAA 68
 DB 1833 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1892
 QY 69 CCAGAAATCCGTTAATATTTGTTAAATATTTACCAATATTTATTAATTTATTTTACG 128
 DB 1893 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1951
 QY 129 CAATAGCATGTTTGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 175
 DB 1952 AATATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2011
 QY 176 CAAAAGTATAAAGTTGATGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 235
 DB 2012 TAAATATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2071
 QY 236 AAGTTGGTTAGACATTTAGCCATTTCTTTTCTTACCAATAAATAAATAAATAAATAA 295
 DB 2072 AAAATATTTTAAATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2131
 QY 296 ACCTTTAGGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 355
 DB 2132 TTTATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2191
 QY 356 ACACCCAGTGACCCACATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 415
 DB 2192 AAAAATAATGAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2251
 QY 416 TTTTATTTCAATACATTTTTCCTACTTCTAAATCTTGGAGATTCACCGTTAACTAATTA 475
 DB 2252 TAAATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2311
 QY 476 GTGTATATCCACGGTCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 535
 DB 2312 TTAATAATTTTGTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2371
 QY 536 CACCTGAACCGCTTATGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 595
 DB 2372 ATAAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2431
 QY 596 AGGCTACTTGGTCTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTG 655
 DB 2432 TTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2491
 QY 656 ACACCGCTACTGAACTATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 715

DB 2492 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2551
 QY 716 ATATCTATATATATATAGGTTTATAGATGCCATTTTAAATAAATAAATAAATAAATAA 775
 DB 2552 ATTTAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2611
 QY 776 ATAGAATAAATAATCGAAAAATGAATGTAATAAATAAATAAATAAATAAATAAATAAATAA 835
 DB 2612 ATATAAATAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2671
 QY 836 AGTCTATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 876
 DB 2672 AATAAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2712

RESULT 4

ABZ10246/c
 ID ABZ10246 standard; DNA, 8056 BP.
 XX AC ABZ10246;
 XX DT 16-JAN-2003 (first entry)
 XX DE Haematopoietic cell proliferation disorder related DNA sequence #386.
 XX KW Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX OS Homo sapiens.
 XX PN WC200277272-A2.
 XX PD 03-OCT-2002.
 XX PF 26-MAR-2002; 2002WO-EP003401.
 XX PR 26-MAR-2001; 2001US-02/9333P.
 XX PA (EPIG-) EPIGENOMICS AG.
 XX PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu B;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
 PI Schwöpe I, Ziebarth H;
 XX WPI; 2003-018942/01.

XX Detecting and differentiating between hematopoietic cell proliferative
 XX disorders, comprises contacting a target nucleic acid with a reagent that
 XX distinguishes between methylated and non-methylated CpG dinucleotides.
 PS Claim 28; SEQ ID NO 386; 117pp; English.

CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of

CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

```

Query Match          5.3%; Score 72.6; DB 7; Length 8056;
Best Local Similarity 44.5%; Pred. No. 0.006;
Matches 425; Conservative 0; Mismatches 519; Indels 11; Gaps 3

```

2	AAATAAATCTATACATTA	AAAAAAATCGATTTTAA	ATTTTAGA	AATTCATG	ATTTATTT	61
QY						
2706	AAATAAATTAATTTTAA	AAAAATTTAATCTTTT	AAAAATTTTAA	ATTTTAA	CA	264
Db						
62	TTTTTACCAGAAATCCG	TTAATTCGTAAATAT	TACC	AAATTTAT	ATAATTTAT	121
QY						
2646	TATTTTCTTATTTT	TTTTTTTAAATAT	TATTTAT	ATAATTTT	TTTTCATATTT	258
Db						

[illegible]

QY 182 TATAAAGTTGATCATGGCGTGGTGGTAGATTATTTTAGTTCTAGGTTCCGAATGCAAGTTG 241

Db 2529 AATTTATTTTTTTTTTATTTTCATTAATATTTTAAATATTTTAAACAAATATATTAATA 247

QY 242 GTTTAGACATTTAGCCCTTATTC^{|||||}TTTCTAACCAAAATAATGTAATGGAAAACCTTT 301

DB 2469 ATATATAAATAAAATAA^{|||||}AAAAAATAATTTT^{|||||}TAATAATAAATAATATTTT^{|||||}TTT 2410

DQ

302 AGGAAAAAAAAAGAAATCAAAAATTGAAGAACATCATCCGGTGGAGTCGAGAGGCCGCACC 361
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db

2409 AATTTCAA AAAATAA TAAAAATT TAATAATTA TTNTAAAA ATATA AAATTA TATATTTAA 2350
|||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

362 ACGTGACCCCAACAATATTAAATAAGAGTTGCTACAGTAATCGGATACTTTTAT 421
2349 AAATTAAATAAATTTATTAAATACAAAAATTAATAATTATTTTAAAAATAAATAAT 229

Oy 422 TCAATCTTTTCCACTCTATAAATCTTGCAGATTGCACTTAATTAAGTGTTA 481
Db 2289 TAAAAATTAAAAATTTTAAATTTTAAATTTTAAATA-----TAAAAAATAAATTTTTT 2233

Qy 482 TATCCACGGTCCTAAANAACATCTGTGTACCGTGCCTCACAATTTCAACATTTGGGACCCCT 541

Db 2236 TTAAATTAACACTTTTTTTTTTTTTTTTTTTTTCATTTTTTTTAAATTTTAAATAAA 2177

[illegible]

QY 602 CTGGCTCTGCTTTGGCTCTTTTGTTCAATTTTTCACGTGATTTTGTGGTGAACACGC 661

Dbb 2116 TATTAAATAAATTATATATTATTATTTATGATTTTAAATAATTTTTTAAATAAAATTTTTTA 2057

Qy 662 GTACTTGAAACCTATTATAAAT..TACATAAATTTTATAAGTTTCACTTCCTATATAACT 720

Dbb 2056 AAAATAAAAAATTTTAAAAATATATTTTATTTTAAATATATTTTAAAAATTTTATTAATAAT 1997

721 CATATATATATAGGGTTTAGATGCCAATTTTAAAAAAAGAAATAAAAAATAATAGA 780
1996 TTAAATAAAAATTAATATAATATTTATTTTAAAAAATTATATTTTTTTTATTTATAT 1937

QY
781 ATAAATTCGAAAAAATGAAGTGTAAAATAATTGAGGGGACCAATAAATATATGAACACTT 840

Db
1936 ATAAATTAAATTAATTTTTTTTATATAAATTTTTATTTTTATATAAAAAAAAAATAAAT 1877

DQ
841 AFTATTAAATTTCCCATAGAACTCAATTCCTTAGTATAAATGAGGTAGCCAGTTGG 900

DB
1876 AATTAATAAATAA TTTTTTTT TTTTTTTTTTTTTTTTTTTTAAAAACAATTTTATTTTAA 1812

901 GAGTACACGAAATGTCA TGAACAGTTGCATGTAGGGAAAATTAAATGTAGTAGA 955
1816 TTATTAAATTAAAAATAAATAACATAATTAAATTTTTTAATTTTAAATTTAAAAA 1762

Ddb

RESULT 5
ABZ10100
ID ABZ10100 standard: DNA: 8056 BP.

XX AC ABZ10100;
XX
DT 16-JAN-2003 (first entry)

XX	Haematopoietic cell proliferation disorder related DNA sequence #240
DE	
XX	Human; haematopoietic cell proliferation disorder; cytostatic;
KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW	cytosine methylation state; gene; ds.

XX OS Homo sapiens.
XX PN WO200277272-A

XX	03-OCT-2002.
PD	
XX	26-MAR-2002;
PF	

XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG

PI Berlin K, Braun A, Di
PI Olek A, Piepenbrock C,
PI Lewin A, Lipscher E,

FI
XX
DR
XX

PT disorders, comprises contacting a target nucleic acid with a reagent that receives and amplifies the signal from the nucleic acid, and performing a PT distinguishes between methylated and non-methylated CpG dinucleotides.

PS Claim 28; SEQ ID NO 240; 117pp; English.

The present invention describes a method for detecting and differentiating between haematopoietic cell proliferative disorders associated with at least 1 gene and/or their regulatory regions in a subject. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. Abz03861 to Abz1118 represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used: for differentiating between healthy haematopoietic cells and proliferative disorder haematopoietic cells; for differentiating between acute lymphocytic leukaemia and acute myelogenous leukaemia; as probes for determining the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) of haematopoietic cell proliferation disorder related sequences and their complements; and as primers for the amplification of haematopoietic cell proliferation disorder related DNA sequences. The nucleotide sequences from the present invention can also be used for detecting a predisposition to, differentiation between subclasses, diagnosis, prognosis, treatment and/or monitoring of haematopoietic cell proliferative disorders. The present method enables highly specific classification of haematopoietic cell proliferative disorders allowing for improved and informed treatment of patients

Query Match	5.2%;	Score 71.4;	DB 7;	Length 8056;
Best Local Similarity	44.4%;	Pred. No. 0.0094;		
Matches 391; Conservative	0;	Mismatches 476;	Indels 14;	Gaps 2

QY 9 TCTATACATTAAAAATTTGATTTTAAAAATTTTAGAAAAATTCATGATTTTATTTTATTTTAA 68

DB 1833 TTTAAAAAAAADAAAAADAAAAADAAAAADAAAAADAAAAADTTAA TTTTAAATAATTTTATTTTATTTTATTTTAA 189

Query Match 4.7%; Score 64.4; DB 6; Length 6106;
 Best Local Similarity 56.5%; Pred. No. 0.13;
 Matches 139; Conservative 0; Mismatches 106; Indels 1; Gaps 1;

QY 1 TAAATATCTATACATTAATAAATTTGATTTTAAATTTTAGAATTCATGATTTTATT 60
 DB 797 TATTTTAAAGTTTTTATTTAGAGTTTAAATTAATAATCGCGTAAATTTTATTTT 856
 QY 61 TTTTACCAGAAATCGGTAAATATTTTAAATATTTACCACTAATTTATAAATTTTA 120
 DB 857 TTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 916
 QY 121 TTTTAAAGCAATTAAGCATGTTT GATAAATATATATATTTTATTTTATTTTCAAA 179
 DB 917 ATTAAATTTGATTTTAAATTTTATTTAGTAAAGATAATATTTTAAAGATTAGATTTGTA 976
 QY 180 AGTATTAAGTTGATGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 239
 DB 977 AATATAAATTAATTAAGAGATTTTGTATTTTATTTTATTTTATTTTATTTT 1036
 QY 240 TGGTTT 245
 DB 1037 GAGTAT 1042

RESULT 9

ABL33223
 ID ABL33223 standard; DNA; 7560 BP.

XX AC ABL33223;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 1196.

XX KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosine; neutropenic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineuritic; antidiabetic; antidiabetic; antipsoriatic;
 KW antineuritic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.

XX OS Homo sapiens.

XX PN WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP007537.

XX PR 30-JUN-2000; 2000DE-01032529.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PA (EP1G-) EPIGENOMICS AG.

XX PI Olek A. Piepenbrock C. Berlin K;

XX DR WPI; 2002-130909/17.

XX PT Nucleic acid comprising fragment of chemically modified gene, useful for
 diagnosis and treatment of diseases associated with abnormal cytosine
 methylation.

XX PS Claim 1; SEQ ID NO 1196; 32pp + Sequence Listing; German.

XX CC The present invention provides a number of human immune system associated
 genes which are modified by the methylation of cytosines. The sequences
 can be used in the diagnosis and treatment of immune system disorders,
 including eye diseases such as retinopathy, neovascular glaucoma and
 macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention

XX SQ Sequence 7560 BP; 2154 A; 47 C; 1554 G; 3805 T; 0 U; 0 Other;

Query Match 4.7%; Score 64.2; DB 6; Length 7560;
 Best Local Similarity 50.5%; Pred. No. 0.14; Indels 0; Gaps 0;
 Matches 156; Conservative 0; Mismatches 153;

QY 1 TAAATATCTATACATTAATAAATTTGATTTTAAATTTTAGAATTCATGATTTTATT 60
 DB 2072 TAAATTAAGTAAGATTTGATTTTATTTTATTTTATTTTATTTTATTTTATT 2131
 QY 61 TTTTACCAGAAATCGGTAAATATTTTAAATATTTACCACTAATTTATAAATTTTA 120
 DB 2132 AATTTTATTTGTAAGTAAATTTTATTTTAAATTTTATTTTATTTTATTTTATT 2191
 QY 121 TTTTAAAGCAATTAAGCATGTTTGAATAAATATATATATTTTAAATTTTCAAAA 180
 DB 2132 GTTTAGATTTATGAAGAATTTTAAAGATTTTAAAGATTTTATTTTATTTAAT 2251
 QY 181 CTATTAAGTTGATGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 240
 DB 2252 TTTTATTTTAAATTTGATATATTTTAAATAATGAGTATTTATTTTATTTTAA 2311
 QY 241 GCTTTAGACATTTACCTTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 300
 DB 2312 GATTTTGAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 2371
 QY 301 TAGGAAAAA 309
 DB 2372 AAAGATAAA 2380

RESULT 10

ABV44994
 ID ABV44994 standard; cDNA; 309 BP.

XX AC ABV44994;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 44985.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219007P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R. Endege WO, Monahan JB;

XX DR WPI; 2001-662795/76.

XX CC Novel isolated nucleic acid molecule associated with cancerous state of
 prostate cells and correlating with presence of prostate cancer, useful
 for detecting presence of prostate cancer, stage of prostate cancer.

XX PS Claim 1; Page 8916; 11750pp; English.

polyglutamine disorder; solid tumour.
 Unidentified.
 WO200192565-A2.
 06-DEC-2001.
 06-APR-2001; 2001WO-EP003973.
 06-APR-2000; 2000DE-01019058.
 07-APR-2000; 2000DE-01019173.
 30-JUN-2000; 2000DE-01032529.
 01-SEP-2000; 2000DE-01043826.
 (EFIG-) EPIGENOMICS AG.
 Olek A, Piepenbrock C, Berlin K;
 WPI; 2002-090046/12.
 New nucleic acids or oligomers, useful for diagnosing or treating
 diseases associated with DNA transcription, e.g. immunological disorders,
 Werner syndrome, psoriasis, myocardial infarction, solid tumors or
 cancer.
 Claim 1; SEQ ID NO 40; 32pp; English.
 The invention relates to a nucleic acid, which comprises a segment of the
 chemically pretreated DNA of genes associated with DNA transcription from
 one of 346 sequences, and an oligomer, in particular an oligonucleotide
 or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 to the chemically pretreated DNA of genes associated with DNA
 transcription. The set of oligomer probes are useful for detecting the
 cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 in a chemically pretreated genomic DNA. The nucleic acids are useful for
 diagnosing or treating diseases associated with DNA transcription
 (particularly with the methylation status), e.g. adenosine deaminase
 deficiency, viral infection, retroviral infection, Sezary syndrome,
 haematological disorders, immunological disorders, Werner syndrome,
 tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 neurological disorders, neurodegenerative disorders, Waardenburg
 syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
 disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 associated genomic DNA molecules of the invention. Note: The sequence
 data for this patent did not form part of the printed specification but
 was obtained in electronic format directly from the European Patent
 Office
 SQ Sequence 7195 BP; 1873 A; 224 C; 1538 G; 3560 T; 0 U; 0 Other;
 Query Match 4.6%; Score 63.4; DB 6; Length 7195;
 Best Local Similarity 50.2%; Pred. No. 0.18; 156; Indels 0; Gaps 0;
 Matches 157; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
 QY 675 ATTAAATTCATTAATTTTATAGTTTCATCTCTTATATATATCTCATATATATATAG 734
 DB 6759 ATTATCCAAAACCAATCATTTTAAATTTTATATTTTAAACATATATATTTTAAAT 6700
 QY 735 CGTTTGAATGCCATTTTAAACAAACAAATTAATTAATTAATTAATTAATTAATTAAT 794
 DB 6699 TATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6640
 QY 795 ATGAATGTAAATTTTGGGGGACAAATTAATTAATTAATTAATTAATTTT 854
 DB 6639 AAACATTTTAAATTTTCAAAAACAAATTAATTTTAAATTAATTAATTAATTAATTAAT 6580
 QY 855 CCATTAGATTTCTATTTTCTTCTAGTTTATATAGTACGCTAGCCAGTTGGAGATACAGAAAA 914
 DB 6579 CCAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 6520

QY 915 TGTCTATGAAACAGTTGCTAGTACGGAATTAATGCTAGTACGCGATAGCAGACAAAAAT 974
 DB 6519 ACTAAGCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6460
 QY 975 CCAAGCCCAAGCTA 987
 DB 6459 ACCCAACCAATA 6447
 RESULT 13
 ABX39419
 ID ABX39419 standard; CDNA; 416 BP.
 XX
 AC ABX39419;
 XX
 DT 20-FEB-2003 (first entry)
 XX
 DE Bovine EST associated with lactation/muscle/fat deposition #4584.
 XX
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW Gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 PN US2002137139-A1.
 XX
 PD 26-SEP-2002.
 XX
 PF 24-SEP-2001; 2001US-00960352.
 XX
 PR 12-JAN-1999; 99JUS-0115707P.
 PR 11-JAN-2000; 2000US-00480902.
 XX
 PA (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;
 WPI; 2003-110599/10.
 XX
 DR New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and analysis,
 PT cattle breeding, or for genetically improving cattle.
 XX
 PS Claim 2; SEQ ID NO 4584; 245pp; English.
 XX
 CC The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived from
 CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
 CC appearing as ABX34836-ABX49947, or complements of them. Also included are
 CC (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non-translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridisation between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
 CC present sequence was not shown in the specification but was obtained in

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 24, 2004, 07:14:56 ; Search time 6100.31 Seconds
(without alignments)
9712.612 Million cell updates/sec

Title: US-09-743-823-21

Perfect score: 1367

Sequence: 1 taaataatctatacattaa.....ctacacacactagaagata 1367

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vl.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo-mus.*

41: em.htgo-other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1367	100.0	1367	6	AX006477	AX006477 Sequence
2	1282.8	93.8	1428	6	AX840232	AX840232 Sequence
3	1282.8	93.8	3393	8	SOY5BG7S	D16107 Glycine max
c 4	82.4	6.0	56152	2	AC116963	AC116963 Dictyoste
5	82.4	6.0	183648	3	AC117076	AC117076 Dictyoste
c 6	81.6	6.0	129240	9	AC084128	AC084128 Homo sapi
7	77.4	5.7	76568	3	MBREV	AF538053 Monosiga
c 8	76.4	5.6	176898	9	AC117569	AC117569 Homo sapi
c 9	76.2	5.6	973	3	AY442221	AY442221 Prosopis
c 10	76.2	5.6	12029	3	AE001427	AE001427 Plasmodiu
c 11	75	5.5	266544	3	AC116956	AC116956 Dictyoste
12	74.6	5.5	8056	6	AX599046	AX599046 Sequence
c 13	74.6	5.5	154071	3	AC115598	AC115598 Dictyoste
14	73.6	5.4	110000	3	AC116305	Continuation (4 of
15	73	5.3	101034	9	AC096550	AC096550 Homo sapi
c 16	72.6	5.3	8056	6	AX599046	AX599046 Sequence
17	72.6	5.3	251448	3	AE014819	AE014819 Plasmodiu
c 18	72.2	5.3	27694	8	HASMT	D31785 Pichia cana
19	72	5.3	254733	3	AC117075	AC117075 Dictyoste
20	71.4	5.2	8056	6	AX598900	AX598900 Sequence
c 21	71.2	5.2	181792	9	AC098822	AC098822 Homo sapi
c 22	71	5.2	191542	2	BX649407	BX649407 Danio rer
23	70.6	5.2	110000	2	PFMAL7P1_00	AL844506 Plasmodiu
24	70.2	5.1	86826	3	PFMAL3B5	AL034556 Plasmodiu
c 25	70.2	5.1	302156	3	AC116977	AC116977 Dictyoste
c 26	69.6	5.1	72243	9	AL731858	AL731858 Human DNA
27	69.4	5.1	252632	3	AE014818	AE014818 Plasmodiu
c 28	69	5.0	250195	3	AE014831	AE014831 Plasmodiu
29	68.8	5.0	163443	2	AC006280	AC006280 Plasmodiu
30	68.8	5.0	205429	2	AC005506	AC005506 Plasmodiu
c 31	68.8	5.0	253132	3	AE014846	AE014846 Plasmodiu
c 32	68.4	5.0	7011	8	DCAJ3139	AE003139 Dausus ca
33	68.2	5.0	57538	2	PFMAL6P1_10	AC115682 Dictyoste
c 34	68.2	5.0	110000	2	PFMAL6P1_10	Continuation (11 o
c 35	68.2	5.0	110000	2	PFMAL7P1_07	Continuation (8 of
36	68.2	5.0	252394	3	AE014833	AE014833 Plasmodiu
37	68	5.0	154995	9	AC011979	AC011979 Homo sapi
c 38	67.8	5.0	6591	8	YSCMTG06	L36890 Saccharomyc
c 39	67.8	5.0	110000	2	PFMAL7P1_06	Continuation (7 of
40	67.8	5.0	348600	1	AB063521	AB063521 Wiggleswo
41	67.4	4.9	12029	3	AE001373	AE001373 Plasmodiu
c 42	67.4	4.9	34996	3	U87145	U87145 Toxoplasma
43	67.4	4.9	132254	3	AC116330	AC116330 Dictyoste
c 44	67.4	4.9	132254	3	AC116330	AC116330 Dictyoste
45	67.4	4.9	253001	3	AE014834	AE014834 Plasmodiu

ALIGNMENTS

RESULT 1	AX006477	Sequence 22 from Patent WO0004146.	1367 bp	DNA	linear	PAT 06-SEP-2000
LOCUS	AX006477					
DEFINITION	AX006477					
ACCESSION	AX006477					
VERSION	AX006477.1	GI:9994604				
KEYWORDS						
SOURCE						
ORGANISM						
		Glycine max (soybean)				
		Glycine max				
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
		rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;				
		Glycine.				
REFERENCE						
AUTHORS		Fogher, C.				

TITLE A synthetic polynucleotide coding for human lactoferrin, vectors,
cells and transgenic plants containing it
JOURNAL Patent: WO 0004146-A 22 JAN-2000;
FOGHER CORRADO (IT); PLANTECHNO SRL (IT)
FEATURES
source 1. 1367
/organism="Glycine max"
/mol_type="unassigned DNA"
/db_xref="taxon:3847"
promoter 1. 1367
/note="sequence of the promoter of the gene coding for 7S
basic globulin"
ORIGIN
Query Match 100.0%; Score 1367; DB 6; Length 1367;
Best Local Similarity 100.0%; Pred. No. 2.6e-202;
Matches 1367; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAAATAATCTATACATTAATAAATTTGATTTTAAATTTTAGAAATTCATGATTTTATT 60
DB 1 TAAATAATCTATACATTAATAAATTTGATTTTAAATTTTAGAAATTCATGATTTTATT 60
QY 61 TTTTATACAGAAATCCGTTTAAATTTGTTTAAATTTATACCAACTAATTTATAAATTTTA 120
DB 61 TTTTATACAGAAATCCGTTTAAATTTGTTTAAATTTATACCAACTAATTTATAAATTTTA 120
QY 121 TTTTAAAGGCAATTAAGCATGTTTGATAAATAATATATATTTGTTATAAATTTTCAAAA 180
DB 121 TTTTAAAGGCAATTAAGCATGTTTGATAAATAATATATATTTGTTATAAATTTTCAAAA 180
QY 181 GTATAAGTTGATGATCGGCTGTTGTTAGTATTTTATTTAGTTTCGATTCGAATGCAAGTT 240
DB 181 GTATAAGTTGATGATCGGCTGTTGTTAGTATTTTATTTAGTTTCGATTCGAATGCAAGTT 240
QY 241 GGTTTAGACATTTAGCCTTTATCTTTTCTTAAACCAAAATAATGTAATGGAACCTT 300
DB 241 GGTTTAGACATTTAGCCTTTATCTTTTCTTAAACCAAAATAATGTAATGGAACCTT 300
QY 301 TAGAANAANAAGAAATCAAAATTTGAACATCATCGGTGGAGTCGAGAAGCCACACC 360
DB 301 TAGAANAANAAGAAATCAAAATTTGAACATCATCGGTGGAGTCGAGAAGCCACACC 360
QY 361 CAGTCACCCCAACAATTAATAAGAGTTTCTCTACAGTAATGCGATACCTTTTATA 420
DB 361 CAGTCACCCCAACAATTAATAAGAGTTTCTCTACAGTAATGCGATACCTTTTATA 420
QY 421 TTCAATACCTTTTCCACTTCTAATAATCTTGGAGATTGACCGTTAACTTAATTAAGTTT 480
DB 421 TTCAATACCTTTTCCACTTCTAATAATCTTGGAGATTGACCGTTAACTTAATTAAGTTT 480
QY 481 ATATCCACCGCTCTAATAAATCTTGTACCGTCTCATTTCACATTTTGGGACCC 540
DB 481 ATATCCACCGCTCTAATAAATCTTGTACCGTCTCATTTCACATTTTGGGACCC 540
QY 541 TGAAGCCGTTATGTTTGGAGTTAGTTTGCACACAGTTGAAGCGATCACTCAGGAGGT 600
DB 541 TGAAGCCGTTATGTTTGGAGTTAGTTTGCACACAGTTGAAGCGATCACTCAGGAGGT 600
QY 601 ACTTGGCTTCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGG 660
DB 601 ACTTGGCTTCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGG 660
QY 661 CGTACTTGAACCTTATTAATAATTTACATTAATTTTAAAGTTTCACTTCTTATATAACT 720
DB 661 CGTACTTGAACCTTATTAATAATTTACATTAATTTTAAAGTTTCACTTCTTATATAACT 720
QY 721 CATATAATATAGGCTTTAGATGCAATTTTAAATAAGATTAATAAATAATTAATAGA 780
DB 721 CATATAATATAGGCTTTAGATGCAATTTTAAATAAGATTAATAAATAATTAATAGA 780
QY 781 ATAAATCGAANAATGAATGTAAATAATTTTGAAGGGGCAAAATTAATATGAAGTCT 840
DB 781 ATAAATCGAANAATGAATGTAAATAATTTTGAAGGGGCAAAATTAATATGAAGTCT 840

QY 841 ATTATTTAAATTTTCCATTAGAAATTTCTATTTTCTTATTTAGTTTAAATAGCTAGCAGTTGG 900
DB 841 ATTATTTAAATTTTCCATTAGAAATTTCTATTTTCTTATTTAGTTTAAATAGCTAGCAGTTGG 900
QY 901 GAGATACACGAAATTTGTCATGAAACAGTTGTCATGAGGAAATTAATGTTAGTAGAGGAT 960
DB 901 GAGATACACGAAATTTGTCATGAAACAGTTGTCATGAGGAAATTAATGTTAGTAGAGGAT 960
QY 961 AGCAAGACAAATTTCCAAAGCCAGCTAGCTCTCAGCGAACTGATCCACACGTCCTTT 1020
DB 961 AGCAAGACAAATTTCCAAAGCCAGCTAGCTCTCAGCGAACTGATCCACACGTCCTTT 1020
QY 1021 ACAGATTTCAAACGGATGAAATTTGATGTCATGCAACTGCAACTAAAGCATTTCTCAGCTGC 1080
DB 1021 ACAGATTTCAAACGGATGAAATTTGATGTCATGCAACTGCAACTAAAGCATTTCTCAGCTGC 1080
QY 1081 CAAGTACCCCTTCAACATCACCACCCCTTTGTTTCTCCCATTTGCATGTTAACTCAAGT 1140
DB 1081 CAAGTACCCCTTCAACATCACCACCCCTTTGTTTCTCCCATTTGCATGTTAACTCAAGT 1140
QY 1141 TTATCCCTTTCTTCTGAAATTTTCAAGCCTTCAAAACAGTCGACGTCCTCAATCTTG 1200
DB 1141 TTATCCCTTTCTTCTGAAATTTTCAAGCCTTCAAAACAGTCGACGTCCTCAATCTTG 1200
QY 1201 TGACCAACAGCCGCAAGAAAGAGATCTCATCCGTTTCACTTAGCCACTTAAAGC 1260
DB 1201 TGACCAACAGCCGCAAGAAAGAGATCTCATCCGTTTCACTTAGCCACTTAAAGC 1260
QY 1261 TAGCCAAACGGTGATCTTCTCTATATATTTGATGCTCTCTAAACAACCAACACTACCAT 1320
DB 1261 TAGCCAAACGGTGATCTTCTCTATATATTTGATGCTCTCTAAACAACCAACACTACCAT 1320
QY 1321 TATTCAATTTCAAACCTTGTCTTATATACACACACTAGAGAATA 1367
DB 1321 TATTCAATTTCAAACCTTGTCTTATATACACACACTAGAGAATA 1367
RESULT 2
AX840232 LOCUS AX840232 1428 bp DNA linear PAT 16-DEC-2003
DEFINITION Sequence 6 from Patent WO03073839.
ACCESSION AX840232
VERSION AX840232.1 GI:39978632
KEYWORDS Glycine max (soybean)
SOURCE Glycine max
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE 1
AUTHORS Fagher, C. and Reggi, S.
TITLE Expression of lysosomal enzymes in plant seeds
JOURNAL Patent: WO 03073839-A 6 12-SEP-2003;
FEATURES Location/Qualifiers
1. 1428
/organism="Glycine max"
/mol_type="unassigned DNA"
/db_xref="taxon:3847"
1. 1428
promoter
ORIGIN
Query Match 93.8%; Score 1282.8; DB 6; Length 1428;
Best Local Similarity 95.6%; Pred. No. 2.8e-189;
Matches 1365; Conservative 0; Mismatches 2; Indels 61; Gaps 2;
QY 1 TAAATAATCTATACATTAATAAATTTTCAATTTTAAATTTTAGAAATTTTATGATTTTATT 60
DB 1 TAAATAATCTATACATTAATAAATTTTCAATTTTAAATTTTAGAAATTTTATGATTTTATT 60
QY 61 TTTTATACAGAAATCCGTTTAAATTTGTTTAAATTTATACCAACTAATTTATAAATTTTA 120

Db 61 TTTTACCAGAAATCGTTAAATATGTTTAAATATTACCACTAATTTATATTTA 120
Qy 121 TTTTAAAGCAATTAAGCAATGTTTGAATAAATATATATATTGTTTAAATATCTTTTCAAAA 180
Db 121 TTTTAAAGCAATTAAGCAATGTTTGAATAAATATATATATTGTTTAAATATCTTTTCAAAA 180
Qy 181 GTATAAAGTTGATGATGCGGTGGTGTAGATATTTTAGTTCTAGGTTTCAAGTCAAGTT 240
Db 181 GTATAAAGTTGATGATGCGGTGGTGTAGATATTTTAGTTCTAGGTTTCAAGTCAAGTT 240
Qy 241 GGTTTAGACATTTACCTTTATCTTTTCTTAAACCAAAATAAATGTTAAATGGAAGCTT 300
Db 241 GGTTTAGACATTTACCTTTATCTTTTCTTAAACCAAAATAAATGTTAAATGGAAGCTT 300
Qy 301 TAGGAAAAAAGAAATCAAAATGAAACATCATCCGGTGGAGTCGAGAGCCACACC 360
Db 301 TAGGAAAAAAGAAATCAAAATGAAACATCATCCGGTGGAGTCGAGAGCCACACC 360
Qy 361 CACGTGACCCACAAATATTAATAAAGAGTTTGTCTACAGTAAATGCGATACTTTTTA 420
Db 361 CACGTGACCCACAAATATTAATAAAGAGTTTGTCTACAGTAAATGCGATACTTTTTA 420
Qy 421 TTCAATATCTTTTCACTTCAAAATCTTGAGATTGCGAGTTTGCACCGTTAACTAATTAAGTT 480
Db 421 TTCAATATCTTTTCACTTCAAAATCTTGAGATTGCGAGTTTGCACCGTTAACTAATTAAGTT 480
Qy 481 ATATCCACGGTCTTAAACCACTTGTACCGTCCCTCACATTTCACTTTGCGACCC 540
Db 481 ATATCCACGGTCTTAAACCACTTGTACCGTCCCTCACATTTCACTTTGCGACCC 540
Qy 541 TGAAGCCGT-TAAGTTAGTTAGTTTGTGAGAGTTGCGAGTTGAGCGCATCACTGAGAGGC 599
Db 541 TGAAGCCGTCTATGTTAGTTAGTTTGTGAGAGTTGCGAGTTGAGCGCATCACTGAGAGGC 600
Qy 600 TAGTGTGCTTGTCTTTCGCTTTTGTTCATTTTTCACGTGATTGTTGTTGTTGACAC 659
Db 601 TAGTGTGCTTGTCTTTCGCTTTTGTTCATTTTTCACGTGATTGTTGTTGTTGACAC 660
Qy 660 GCGTACTTGAACTTATTAATAATACATAATTTTAAAGTTTCACTTTTATATATATAC 719
Db 661 GCGTACTTGAACTTATTAATAATACATAATTTTAAAGTTTCACTTTTATATATATAC 720
Qy 720 ----- 719
Db 721 TTCATTCATGCAATTTATTAATTTTGAATGAATAATAAGAGTTTGTAAAAAATATATTAT 780
Qy 720 TCATATAATATATAGGTTTGAATGCGCAATTTTAAAGAAAGATAAATAAATAATATAG 779
Db 781 TCATATAATATATAGGTTTGAATGCGCAATTTTAAAGAAAGATAAATAAATAATATAG 840
Qy 780 AATAAATCGAAAAATGAAATGTAATAAATTTGAGGGGGAACAATAAATATGAAGTC 839
Db 841 AATAAATCGAAAAATGAAATGTAATAAATTTGAGGGGGAACAATAAATATGAAGTC 900
Qy 840 TATTATTATATTTTCCATTTAGATTTCTATTTCCTTAGTTAAATAGAGCTAGCAGTTG 899
Db 901 TATTATTATATTTTCCATTTAGATTTCTATTTCCTTAGTTAAATAGAGCTAGCAGTTG 960
Qy 900 GGAGATACGAAAAATGTCATGAACAGTTGCAATGAGGGAATTAATGTAGTAGAGGA 959
Db 961 GGAGATACGAAAAATGTCATGAACAGTTGCAATGAGGGAATTAATGTAGTAGAGGA 1020
Qy 960 TAGCAAGCAAAATTCGAGCAAGCTAGCTGCTCAGCGGAACTGATCCACAGTCTT 1019
Db 1021 TAGCAAGCAAAATTCGAGCAAGCTAGCTGCTCAGCGGAACTGATCCACAGTCTT 1080
Qy 1020 TACAGAGTTTCAAAAGGATGAATCTGATGCAATGCACTGAACATTTGTTCTCAGCTG 1079
Db 1081 TACAGAGTTTCAAAAGGATGAATCTGATGCAATGCACTGAACATTTGTTCTCAGCTG 1140
Qy 1080 CCAAGTACCCCTCAACCTCAACCAACCTTTGTTTCTCCCATGATGATTAAGTCAAG 1139
Db 1141 CCAAGTACCCCTCAACCTCAACCAACCTTTGTTTCTCCCATGATGATTAAGTCAAG 1200

Qy 1140 TTTATCTTCTTTCTTCTTCTGGAATTTTCAAGCCTCAAAACAGTCGAGCTCAATCTT 1199
Db 1201 TTTATCTTCTTCTTCTTCTGGAATTTTCAAGCCTCAAAACAGTCGAGCTCAATCTT 1260
Qy 1200 GTCAACCAACAGCGCCCAAAAGAAAGAAATCTCATCCCGTTACACTTAGCCACTTAAAG 1259
Db 1261 GTCAACCAACAGCGCCCAAAAGAAAGAAATCTCATCCCGTTACACTTAGCCACTTAAAG 1320
Qy 1260 CTAGCCAAACGGTGATCTTCTCTATATATTTAGTCTCTTAACAACAACACTATCA 1319
Db 1321 CTAGCCAAACGGTGATCTTCTCTATATATTTAGTCTCTTAACAACAACACTATCA 1380
Qy 1320 TTTATCAATTTCAAACTTGTCTTCTATCTACACACTAGAGAATA 1367
Db 1381 TTTATCAATTTCAAACTTGTCTTCTATCTACACACTAGAGAATA 1428

RESULT 3
SOYSEB7S
LOCUS 3393 bp DNA linear PLN 01-AUG-2002
DEFINITION Glycine max gene for basic 7S globulin, complete cds.
ACCESSION D16107
VERSION D16107.1 GI:434060
KEYWORDS
SOURCE Glycine max (soybean)
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1
AUTHORS Kagawa, H. and Hirano, H.
TITLE Sequence of a cDNA encoding soybean basic 7S globulin
JOURNAL Nucleic Acids Res. 17 (21), 8868 (1989)
MEDLINE 90067863
PubMed 2587227
2
AUTHORS Watanabe, Y. and Hirano, H.
TITLE Nucleotide sequence of the basic 7S globulin gene from soybean
JOURNAL Plant Physiol. 105 (3), 1019-1020 (1994)
MEDLINE 94336768
PubMed 8058830

REFERENCE 3 (bases 1 to 3393)
AUTHORS Watanabe, Y.
TITLE Cloning and sequence analysis of the basic 7S globulin gene in soybean
JOURNAL Unpublished
REFERENCE 4 (bases 1 to 3393)
AUTHORS Watanabe, Y.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1993) Yoshihiro Watanabe, National Institute of Agrobiological Resources, Department of Molecular Biology, 2-1-2 Kannondai, Tsukuba, Ibaraki 305, Japan (Tel:81-298-38-7031, Fax:81-298-38-7032)

FEATURES
source Location/Qualifiers
1..3393
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Miyagishirone"
/db_xref="taxon:3847"
/clone="pBGW1 and pBGW2"
/tissue_type="leaf, etiolated"
/clone_lib="pSOY-418"
/dev_stage="seedling"
1216..1230
/standard_name="heat shock like element"
/notes="HSE"
/evidence=not_experimental
1254..1257
TATA_signal
1344..1350
misc_feature
/note="primary transcription start site"

misc_signal
1216..1230
/standard_name="heat shock like element"
/notes="HSE"
/evidence=not_experimental
1254..1257
TATA_signal
1344..1350
misc_feature
/note="primary transcription start site"

gene	1429. .2712	QY	600	TACTGTCTTTCCTTTTCGCTTTTGTCAATTTTCACTGATTTTGTGTGGGAACAC	659
CDS	/gene="SBg7S"	DB	601	TACTGTCTTTCCTTTTCGCTTTTGTCAATTTTCACTGATTTTGTGTGGGAACAC	660
	/note="Bg"	QY	660	CGTACTTGAACCTTATTAATAATACATAATTTTATAAGTTTCACCTCTTATATATAC	719
	/codon_start=1	DB	661	CGTACTTGAACCTTATTAATAATACATAATTTTATAAGTTTCACCTCTTATATATAC	720
	/evidence=not experimental	QY	720	-----	719
	/product="basic 7S globulin"	DB	721	TTCAATTCATGATTTAATTTTTCATCAATAATAAGAGTTTGTATAAATAATATATT	780
	/protein_id="BAA03681.1"	QY	720	TCATAATAATATAGGTTTAGAATGCAATTTTATAAAGAGTAATAAATAATAATAG	779
	/db_xref="GI:434061"	DB	781	TCATAATAATATAGGTTTAGAATGCAATTTTATAAAGAGTAATAAATAATAATAG	840
	/translation="MASILHVLALSLSCSLFPLSDSVTPKPLNLVLPVNDGST	QY	780	AATAAATCGAAAAATGAAATGTAAAAAATTTGAGGGGCAAAATAAATAATAGAAATC	839
	GLHWAQKRPPLMQPVLDVINGHLWNCEQOISSKYQAPPHSTOCSSRANTHQ	DB	841	AATAAATCGAAAAATGAAATGTAAAAAATTTGAGGGGCAAAATAAATAATAGAAATC	900
	LSCPAARSGCHNCTGLMTPITQOTGLGELGEDVGAHATQGSTQGLGELVTPQC	QY	840	TATTAATAATTTTCCATTTAGAAATTTCTATTTTCTTAGTTAATATAGGTAGCCATG	899
	FLFCAPSLVQLGRLPNTQVAGLGHAPISLQNLASHFLGROFTTCLISRYPTSKG	DB	901	TATTAATAATTTTCCATTTAGAAATTTCTATTTTCTTAGTTAATATAGGTAGCCATG	960
	AIIFGAPNMRQFQDIFHDLAFTPLITLQGEYNVRVNSIRINQHSVFPLNKS	QY	900	GGAGATACAGAAAAATGTCATGAACAGTTGCATGTAGGGGAAATTAATGTAGTAGAGGA	959
	TIIVGSTSGMTISSTPHVLOQSVYQQLPKQAVKSVAPFLGFCFNKI	DB	961	GGAGATACAGAAAAATGTCATGAACAGTTGCATGTAGGGGAAATTAATGTAGTAGAGGA	1020
	NAVPSVDLVMDKPGVPWIRISGDLNVQAPQVTCIGVWNGNQPRABETILGARLEE	QY	960	TAGCAGACAAAAATCCAAGCCAGTAGCTGCTCAGCGGAACCTCGATCCACAGCTCTT	1019
	NLVVPLARSRVGFSSLSHSHGVKCADLPENFA"	DB	1021	TAGCAGACAAAAATCCAAGCCAGTAGCTGCTCAGCGGAACCTCGATCCACAGCTCTT	1080
sig_peptide	/gene="SBg7S"	QY	1020	TACAGAGTTTCAACCGGATGAATCTGCATGGCACTCAACTAAGCAATTTGTTCTCAGCTG	1079
	/citation=[1]	DB	1081	TACAGAGTTTCAACCGGATGAATCTGCATGGCACTCAACTAAGCAATTTGTTCTCAGCTG	1140
mac_peptide	/gene="SBg7S"	QY	1080	CCAAGTACCCCTCACACTCACCAACCCCTTTGTTTCTCCCAATTCGATGTTAACTCAAG	1139
	/product="basic 7S globulin"	DB	1141	CCAAGTACCCCTCACACTCACCAACCCCTTTGTTTCTCCCAATTCGATGTTAACTCAAG	1200
	/citation=[1]	QY	1140	TTTATCTTTCTTTCGCTTCTGAAAATTTTCAAGCCCTCAACAGCTCGAGCTCAATCTT	1199
polyA_signal	2786. .2791	DB	1201	TTTATCTTTCTTTCGCTTCTGAAAATTTTCAAGCCCTCAACAGCTCGAGCTCAATCTT	1260
polyA_signal	2806. .2811	QY	1200	GTGACCAACAGCGGCAAAAGAGACATCTCATCCCTTCACTTACCTTAAAG	1259
ORIGIN		DB	1261	GTGACCAACAGCGGCAAAAGAGACATCTCATCCCTTCACTTACCTTAAAG	1320
	Query Match 93.8%; Score 1282.8; DB 8; Length 3393;	QY	1260	CTAGCCAAACGGTGATCTTTCTTATATATTTAGTCTCTTAACACCAACCACTACCA	1319
	Best Local Similarity 95.6%; Pred. No. 2.4e-189;	DB	1321	CTAGCCAAACGGTGATCTTTCTTATATATTTAGTCTCTTAACACCAACCACTACCA	1380
	Matches 1365; Conservative 0; Mismatches 2; Indels 61; Gaps 2;	QY	1320	TTATTCAATTTCAACCTTCTCTCTATACACACTAGACATA 1367	
		DB	1381	TTATTCAATTTCAACCTTCTCTCTATACACACTAGACATA 1428	
		RESULT 4			
		AC116963/c			
		LOCUS	AC116963	56152 bp	DNA linear HTG 04-APR-2002
		DEFINITION	Dictyostelium discoideum chromosome 2 map 4790445-4846595 strain		
		ACCESSION	AC116963		
		VERSION	AC116963.1		
		KEYWORDS	HTG; HTGS PHASE2		
		SOURCE	Dictyostelium discoideum		
		ORGANISM	Dictyostelium discoideum		
		REFERENCE	1 (bases 1 to 56152)		
		AUTHORS	Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.		
			Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P.,		
			Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kumpf, K.,		
			Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and		
			Noegel, A. A.		

Sequence and Analysis of Chromosome 2 of Dictyostelium
 Unpublished
 The Dictyostelium Genome Sequencing Consortium
 2 (bases 1 to 56152)
 Baumgart, C.
 Direct Submission
 Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 CDS predictions from GenBank may contain errors. Further information
 is available from IMB Jena, Department of Genome Analysis
 (<http://genome.imb-jena.de/dictyostelium/>)
 and the University Cologne, Institute for Biochemistry I
 (<http://www.uni-koeln.de/dictyostelium/project.shtml>)
 Funding
 Agency : Deutsche Forschungsgemeinschaft (DFG).
 * NOTE: This is a "working draft" sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 Location/Qualifiers
 1..56152
 /organism="Dictyostelium discoideum"
 /mol_type="genomic DNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /chromosome="2"
 /map="4790445-4846595"

ORIGIN

Query Match 6.0%; Score 82.4; DB 2; Length 56152;
 Best Local Similarity 44.3%; Pred. NO. 0.0011;
 Matches 430; Conservative 0; Mismatches 531; Indels 9; Gaps 2;

QY 13 TACATTAATAATTTGATTAAATTTTAAATTTTACCACTAATTTATTAATTTTATTTTATTTTACAG 72
 Db
 QY 36134 TACAAGATGAATTTGTTTATTTGCAATGAAATGAAATTAATAATCAACACACAC 36075
 Db
 QY 73 AAATCCGTTAATTTGTTAAATATTTACCACTAATTTATTAATTTTATTTTAAAGCAAT 132
 Db
 QY 36074 AAATTTGTTGAATTAATCATCATTTTCAAAATTTGTTATATTTGTTATATAAAAAA 36015
 Db
 QY 133 TAAGCATGTTTCATTAATAATATATTTGTTATTAATTTTCTAAAGTATTAAGTTGA 192
 Db
 QY 36014 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 35955
 QY 193 TGATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 252
 Db
 QY 35954 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 35895
 QY 253 TAGCCTTATCTTTTCTTACCAATAATAATGTAATGGAATGGAATGGAATGGAATGGA 312
 Db
 QY 35894 ATATATATTTATATGTCATATTAATAATAATAATAATAATAATAATAATAATAATA 35835
 QY 313 GAATCAAAATGAAACATCATCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 372
 Db
 QY 35834 AATACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 35775
 QY 373 CAATATTAATAATGAGTTTGGTCTACAGTAATGCGATACCTTTTATTAATCAATCTTT 432
 Db
 QY 35774 TAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 35715
 QY 433 TCACCTCTAAATCTGGAGATTTGACCGGTACTACTAATTAAGTTATATCCACGGT 492
 Db
 QY 35714 TTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 35655
 QY 493 CTTAAAAAATCTGTGTACCGTCTCAATTTCACTTTGGCCACCTGGAAGCGTTA 552
 Db
 QY 35654 ATAAATGACATGAAATATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 35595
 QY 553 TGTTAGGTAGTTTGGACAGTTGAAGCGCATCACTCAGGAGGCTACTTGGTCTTGC 612
 Db
 QY 35594 AAATCTAGGTTCTATATATCAAAAAAATAAATCAATTTATTAATTTTATTTTATTT 35535

QY 613 TTTTGGTCTTTTGTTCATTTTTCAG-----TGATTTGTTGGTGAACGGGTAC 665
 Db
 QY 35534 TTTTGT 35475
 QY 666 TTGAACCTTATTATAAATTACATAAATTATTAAGTTTCTTCTTATATAAATCTCATAT 725
 Db
 QY 35474 TTTTGT 35415
 QY 726 AATATATAGGTGTAGAG--CCAAATTTTAAAAAGAAATAAAAAATAAATAGAAATA 783
 Db
 QY 35414 AATATAAATAATTAATAAAAAAATAAATAAAAAAATAAATAAAAAAATAAATAAAAA 35355
 QY 784 AATCGAAAAAATGAATGTAAAAAATTTGAGGGGACAAATTAATAATATGAAGTCTATT 843
 Db
 QY 35354 AAAAAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 35295
 QY 844 ATTTAAATTTTCCATAGAAATCTATTTTCTTGTAGTTATATGAGCTAGCAGCTGGAG 903
 Db
 QY 35294 TTAACAGATTAATAAGAAATTCATCCCTAAAAAGAAAAATAAATAAAAAAATA 35235
 QY 904 ATACACGAAATCTCATGAACAGTTTGCATGTAGGAAATTAATCTAGTAGGGATAGC 963
 Db
 QY 35234 AATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 35175
 QY 964 AAGACAAAAA 973
 Db
 QY 35174 AGTGTAATAA 35165

RESULT 5
 AC117076 183648 bp DNA linear INV 12-MAR-2003
 LOCUS Dictyostelium discoideum chromosome 2 map 5862124-6045772 strain
 DEFINITION AX4, complete sequence.

AC117076 GI:28829496
 AC117076
 HTG.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryote; Mycetozoa; Dictyostelida; Dictyostelium.
 REFERENCE 1 (bases 1 to 183648)
 Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P.,
 Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kump, K.,
 Tungal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
 Noegel, A. A.
 Sequence and analysis of chromosome 2 of Dictyostelium discoideum
 Nature 418 (6893), 79-85 (2002)

TITLE Nature 418 (6893), 79-85 (2002)
 JOURNAL Nature 418 (6893), 79-85 (2002)
 MEDLINE 22092622
 PUBMED 12097910
 REMARK The Dictyostelium Genome Sequencing Consortium
 REFERENCE 2 (bases 1 to 183648)
 Baumgart, C.
 DIRECT SUBMISSION
 Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

TITLE Direct Submission
 JOURNAL Direct Submission
 REFERENCE 3 (bases 1 to 183648)
 Baumgart, C.
 DIRECT SUBMISSION
 Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 REFERENCE 4 (bases 1 to 183648)
 Baumgart, C.
 DIRECT SUBMISSION
 Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

COMMENT On Mar 4, 2003 this sequence version replaced gi:20066249.
 CDS predictions from GenBank do not necessarily reflect true genes.
 Further information is available from IMB Jena, Department of
 Genome Analysis
 (<http://genome.imb-jena.de/dictyostelium/>)
 and the University Cologne, Institute for Biochemistry I
 (<http://www.uni-koeln.de/dictyostelium/project.shtml>)
 Funding

ALQKLAKHLVNEKRLDLCKRILKILFQOLGKSPRRGDSNFSNTNGGGGGGCGNFSN
 YNMDPNVLSPIKXMGFLPILTAELRTTPIRLDSNSGWNFNAVPPPTGINSDE
 BLFNLSADSVKLVSVAFYCVPRKPSIGQWRSDWLSVOMIAHLSPLISNPKPHY
 LEMEVELDLILKQLEKDYFVNKQSLVEQIPNTAYIKNLTLAQGLFHSWILEM
 RPAVSGSFELAFGLYLEDGLENINVSICIRALVDKVPADFLTGHPSTPIAYELS
 LHMDFLLFCYFPEPRKADDDYILSTLVKPPQVWSKGLCLSTLLDIEAIGKGA
 RMDMISVRLTHSHLILPELTPVRQLGLDILALGNILKSGTNRAPSEIQLBLE
 YMQRFNQLKHQHMGLSLAVEIGSNADLAADKLISNHGAELITNQAAIINNOQFNQSG
 GLASTCTGNAANFVNSLQKLVLYGGEIKMLHMNMVDERVQOKLFOQOQKQNGED
 NDDDDDDDEBTSEIEDRAVDLIVKEFENLISNVKSNHPNANMARTSIMYRSCAPL
 INHYEGSNVRLTHMICYAPANTPTKESMEIGUCWKLIAERPDLSPIMTSISDIWT
 WTVNORIGLFPANNREPSPLSLSQOTSQDNDKQVNEILSTSPNNNNNNNNNNNN

Query Match 6.0%; Score 82.4; DE 3; Length 183648;
 Best Local Similarity 44.3%; Pred. No. 0.00069;
 Matches 430; Conservative 0; Mismatches 531; Indels 9; Gaps 2;

QY 13 TACATTAATAAATTTGATTTTAAATTTTAGAATTCATGATTTTATTTTATACAG 72
 DB |||||
 QY 34352 TACACATGAATTTGTTTATTGCAAAATGAATTAATAAATAATCAACACACAC 34411
 DB |||||
 QY 73 AAATCGGTTAATTTGTTAAATAATATACCACTAATTTATTAATTTTATTTTAAAGCAAT 132
 DB |||||
 QY 34412 AAATGTTTGAATTAATCATCATTTTCAAAATTTTGATATATTGTATATAAAAAA 34471
 DB |||||
 QY 133 TAAGCATGTTTCATAAATAATATATATGTTTATAATCTTTTCAAAAGTATTAAGTTGA 192
 DB |||||
 QY 34472 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 34531
 DB |||||
 QY 193 TGATGCGTGGTGGTAGATTATTATTAGTCTAGGTGCAATGCAATGTTGGTTAGACATT 252
 DB |||||
 QY 34532 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 34591
 DB |||||
 QY 253 TAGCCTATCTTTTCTTACCAATAAATATGTAATGGAACACTTTTAGGAATAA 312
 DB |||||
 QY 34592 ATATATATTATTATGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 34651
 DB |||||
 QY 313 GAATCAAAATTCAGAACATCTCCGGTGGAGTGCAGAGCCACACCCACCGTCCCAA 372
 DB |||||
 QY 34652 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 34711
 DB |||||
 QY 373 CAATATAAATAAAGATTTGCTACAGTAATGCAATCTTCACTTTGCGCACCTGGAAGCGTTA 432
 DB |||||
 QY 34712 TAAAGTTGGTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTATTA 34771
 DB |||||
 QY 433 TCACCTTCATAATCTTGAGATTTGCACCGTAACTAATTAAGTTGTTATATCCACCGT 492
 DB |||||
 QY 34772 TTTATTTATTTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAAGCA 34831
 DB |||||
 QY 493 CCTAAAAAATCTGTACCGTGCCTCACATTTCACTTTGCGCACCTGGAAGCGTTA 552
 DB |||||
 QY 34832 ATAAATGACGAAATATTCATCAATTTTATTTTAAAAAATAAAGTGATTTT 34891
 DB |||||
 QY 553 TGTTAGGTAGTTGTTTGGCAACAGTTGAAGCGCATCACTCAGAGCGTACTTGGCTTGC 612
 DB |||||
 QY 34892 AAATCAGGTTCTATATATCAAAAAAATAAATAAATAAATAAATAAATAAATAA 34951
 DB |||||
 QY 613 TTTTGGCTCTTGTCTCAATTTTTCAG-----TGATTTGTTGGTGAACACGGTAC 665
 DB |||||
 QY 34952 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 35011
 DB |||||
 QY 666 TTGAACCTTATTAATAATACATAATTTTATAAGTTTCACTTCTTATATAAATCAATAT 725
 DB |||||
 QY 35012 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 35071
 DB |||||
 QY 726 AATATATAGGTTTAGAATG--CCATTTTTHAAAAAATAAATAAATAAATAAATAA 783
 DB |||||
 QY 35072 AATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 35131
 DB |||||
 QY 784 AAATCAAAAAATGAATGTAAAAAATTTGAGGGGACAAATAAATAAATAAATAAATAA 843
 DB |||||
 QY 35132 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 35191
 DB |||||

QY 844 ATTTAAATTTTCCATTAGAAATTCATTTTCTTTAGTTTAATATATGAGCTAGCAGCTGGAG 903
 DB 35192 TTAACACAGATTATTAAGAAATTCATCCCTAAAAAAGAAAAATAAATAAATAAATA 35251
 QY 904 ATACACAAATATGTCATGAACAGTTCATCTAGTAGGGAATTAATGTAGTACAGGGTAGC 963
 DB 35252 AAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 35311
 QY 964 AAGACAAAAA 973
 DB 35312 AGTGTAAAAA 35321

RESULT 6 AC084128/c

LOCUS Homo sapiens chromosome 8, clone CTD-3118D11, complete sequence.
 DEFINITION Homo sapiens chromosome 8, clone CTD-3118D11, complete sequence.
 ACCESSION AC084128
 VERSION AC084128.6 GI:22004605
 HTG.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 129240)
 Birren, B., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 8, clone CTD-3118D11
 Unpublished
 2 (bases 1 to 129240)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Beda, P., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
 Choepel, I., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferrera, P.,
 FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
 Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
 Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
 MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
 McPheters, R., Meldrum, J., Meneus, L., Minova, T., Mlenga, V.,
 Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
 O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K.,
 Pierron, P., Pisani, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
 Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
 Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
 Zimmer, A. and Zody, M.

Direct Submission
 Submitted (13-OCT-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 129240)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Paro, S., Ferreira, P., FitzGerald, M., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Gird, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
 Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,
 Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
 Meneus, L., Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
 Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,
 Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
 Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (02-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 129240)
 Biren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barua, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Canarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dekrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneses, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 30, 2002 this sequence version replaced gi:21306731.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 3118D11
 Center clone name: 3118_D11

----- Location/Qualifiers
 1..129240
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="CTD-3118D11"
 /clone_lib="CTD2 Human BAC"
 1..224
 /rpt_family="AluSg/x"
 225..403
 /rpt_family="AluJb"
 669..855
 /rpt_family="L1MC3"
 1519..1600
 /rpt_family="AT rich"
 complement(1619..1895)
 /rpt_family="AluSc"
 complement(1897..2197)
 /rpt_family="AluSg"
 2310..2593
 /rpt_family="Char11c2"
 complement(2598..3451)
 /rpt_family="L1MA4A"
 2966..2969
 /note="single clone coverage"
 3483..3713
 /rpt_family="L1MDa"
 complement(3714..4017)
 /rpt_family="AluSg"
 3739..3748
 /note="<30 qual SNGL region"

unsure
 3785..3811
 /note="<30 qual SNGL region"
 4018..4255
 /rpt_family="L1MDa"
 4267..4982
 /rpt_family="L1MDa"
 4974..5464
 /rpt_family="L1MD2"
 5827..6015
 /rpt_family="MIR"
 8136..8345
 /rpt_family="MIR"
 8552..8737
 /rpt_family="L2"
 9347..9644
 /rpt_family="AluSx"
 complement(10188..10324)
 /rpt_family="MIR3"
 10770..11002
 /rpt_family="MIR"
 complement(11116..11638)
 /rpt_family="MER1A"
 12153..12222
 /rpt_family="L2"
 complement(12568..12878)
 /rpt_family="MIR"
 complement(13676..14221)
 /rpt_family="MER41A"
 complement(15056..15245)
 /rpt_family="AluJb"
 15272..15714
 /rpt_family="AluSx"
 complement(16162..16386)
 /rpt_family="MIR"
 16712..17008
 /rpt_family="AluJo"
 complement(18862..18963)
 /rpt_family="MERSA"
 complement(19028..19089)
 /rpt_family="L2"
 19090..19375
 /rpt_family="AluJb"
 complement(19376..19430)
 /rpt_family="L2"
 complement(19480..19643)
 /rpt_family="L2"
 complement(19744..19829)
 /rpt_family="MIR3"
 complement(19976..20109)
 /rpt_family="MIR"
 22048..22074
 /rpt_family="(T)n"
 complement(22240..22864)
 /rpt_family="L1PA7"
 complement(23650..23959)
 /rpt_family="AluJb"
 complement(24055..24374)
 /rpt_family="AluSg1"
 24894..24922
 /rpt_family="AT rich"
 complement(24934..25223)
 /rpt_family="AluJb"
 26058..26354
 /rpt_family="AluSx"
 26936..27229
 /rpt_family="AluSx"
 28099..28243
 /rpt_family="AluSx"
 29591..29611

Query Match
 Best Local Similarity 44.1%; Pred. No. 0.0013;
 Matches 383; Conservative 0; Mismatches 484; Indels 1; Gaps 1;
 6.0%; Score 81.6; DB 9; Length 129240;

QY 736 GTTTAGATGCCAATTTTAAAAAAGATATAAAAAAATAATAGATAATATAATCGAAAAA 795
 Db 112096 ATATATTATATATATAATATAATATAATATAATATAATATAATATAATATAATATA 112037
 QY 796 TGAATGTAAAAAATTTGAGGGGCAATAAAAAATGAAAGCTATATTATTTAAATTTTC 855
 Db 112036 ATATATTATATATATATTTATATAATATAATATAATATAATATAATATAATATAAT 111977
 QY 856 CATTAGATCTATTTTCCTTAGTTA 881
 Db 111976 GATATATATATATTTTATATATTA 111951

RESULT 9
 AY442221/c
 LOCUS 973 bp DNA linear INV 18-NOV-2003
 DEFINITION Drosophila subobscura strain Boquillobo-32 A-T rich region genomic
 sequence; mitochondrial.
 ACCESSION AY442221
 VERSION AY442221.1 GI:38304739
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila subobscura
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 973)
 AUTHORS Brehm, A., Harris, J., Hernandez, M., Perez, J., Larruga, J., Pinto, F.
 and Gonzalez, A.
 TITLE Phylogeography of *Drosophila subobscura* from North Atlantic Islands
 inferred from mtDNA A+T rich region sequences
 JOURNAL Mol. Phylogenet. Evol. (2003) In press
 REFERENCE 2 (bases 1 to 973)
 AUTHORS Brehm, A.
 TITLE Direct Submission
 JOURNAL Submitted (20-OCT-2003) Centre of Macaronesian Studies, University
 of Madeira, Campus of Penteada, 9000 Funchal, Portugal
 FEATURES
 source
 1..973
 /organism="Drosophila subobscura"
 /organelle="mitochondrion"
 /mol_type="genomic DNA"
 /strain="Boquillobo-32"
 /db_xref="taxon:7241"
 /country="Portugal; Paul Boquillobo"
 misc_feature 1..973
 /note="A+T rich region"

ORIGIN
 Query Match 5.6%; Score 76.2; DB 3; Length 973;
 Best Local Similarity 44.7%; Pred. No. 0.021;
 Matches 385; Conservative 0; Mismatches 469; Indels 8; Gaps 2;
 QY 1 TAAAAATCTATACATATAAAAAATTTGATTTTAAAAATTTAGAAATTCATGATTTATT 60
 Db 941 TTAATATATATATAATATAATATAATATATGATAAATAATTTAAAAATTTAAAT 882
 QY 61 TTTTTCACCAAGATCCGTTAATTTGTTAAATATATACCAACTTAATTTAAATTTA 120
 Db 881 TTATATTAATATAACAAAAATAAATTTTATTTAATTAATCTGATAAATAATAA 822
 QY 121 TTTTAAAGCAATAGCATGTTTGAATAAATAATATATTTGTTATAATATCTTTCAAA 180
 Db 821 TAATAAATAAATAATAAATTTATTTTAAATAAATAATTTTAAATCCCTCGAATAT 762
 QY 181 GTATAAAGTCATGATGGCGTGGTGAATTTTATTTAGTTCTAGTTCCGAATCGAAGTT 240
 Db 761 AAATTTTATTTATAAAAATCTTTATTTTATTAATAAGAAAGAGATTTTAAAAACTTT 702
 QY 241 GGTTAGACATTTACCTTATCTTTTCTTAA CCAAAATAAATGTAATGGAACCTT 300
 Db 701 TTTATATTATTAACCTTATAAATTTTATTAATAAATAATAATTTTAAATGAATATTTT 642

QY 301 TAGAAAAAAGAAATCAAAATTTGAAAAATCATCATCCGGTGGAGTCGAGAGCCACACC 360
 Db 641 TAATAAATAAATAAATTTAGTAATTTAAACAATAATATATATATATATATATATATA 582
 QY 361 CACGTGACCCCAACAATATTAAATAAAGAGTTTGTCTACAGTAAGTAATCGGATACCTTTT 420
 Db 581 TTTATTTTAAATTTAAATTTAAATAATATATATTAATAATAATCAAAATAATTTATAT 522
 QY 421 TTCAATACCTTTTCCACTTCTAAATCTTGGAGATTTGCAACCGTTAACTAATTAAGTGT 480
 Db 521 TTATATA-TATATAAATAATATACCATATAATTAATTAATTAATTAATAAATAATAGTA 463
 QY 481 ATATCCACCGCTCTAAAAAAGCTGTGTACCGCTGCACATTCACATTTGCGGACCC 540
 Db 462 ATTTATATATATATATAAATAATTTACTATTTTATTTTATATATACTAGTTCTAGATTA 403
 QY 541 TGAAGCCGTTATGTTTGGTGTAGTTGTTTCAACAGTTGAAGCGCATCACTCAGGAGGCT 600
 Db 402 ATAGATAATATATATATATAAATGTTTAAATATATATCTTATATAGTTAACAATTTAAAT 343
 QY 601 ACTTGGTCTTGTCTTCCGCTTTTGTTCATTTTTCACGTGATTTTGTGGTGGAACAG 660
 Db 342 AAAAAATTTCCAGTCCAAAAAATGGTGACATAATATAAATAAGTTATTAGTTAATAAT 283
 QY 661 CGTACTTGAA-----ACTTATTATAAATACATAATTTTATAAGTTTCCACTTCTATA 713
 Db 282 TTATATAACATCTATCTATTATATAAATAATATAAATTTTATTAATTTATATATTT 223
 QY 714 TAATACATATATAATATATAGGTTTAGAATGCCAATTTTAAAAAAGAAATAAAAAAT 773
 Db 222 ATATATTATATATAGTTGATTAATTTTATATAACATATATATATATAGAAAAATAAAT 163
 QY 774 AATAGATTAATTCGAAAAAATCGAATGTGAAAAAATTTGAGGGGCAATAAATAATG 833
 Db 162 TATTTAATTAATAAATAATTTTATAAATAAATAATTTACTAATTTCTATTATTATAATA 103
 QY 834 AAGTCTATTATTATAAATTT 854
 Db 102 AATTAAATTTTATAAATCT 82

RESULT 10
 AE001427/c
 LOCUS 12029 bp DNA linear INV 04-OCT-2002
 DEFINITION Plasmodium falciparum 3D7 chromosome 2 section 64 of 73 of the
 complete sequence.
 ACCESSION AE001427 AE001362.
 VERSION AE001427.1 GI:3845314
 KEYWORDS
 SOURCE
 ORGANISM
 Plasmodium falciparum 3D7
 Plasmodium falciparum 3D7
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 12029)
 AUTHORS Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravid, L.,
 Kocun, E.V., Shaloom, S., Mason, T., Yu, K., Fujii, C., Pedersen, J.,
 Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Perle, M.,
 Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O.,
 Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C. and Hoffman, S.L.
 Chromosome 2 sequence of the human malaria parasite Plasmodium
 falciparum
 TITLE
 JOURNAL Science 282 (5391), 1126-1132 (1998)
 MEDLINE 99021743
 PUBMED 9804551
 REFERENCE 2 (bases 1 to 12029)
 AUTHORS Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W.,
 Carlton, J.M., Paul, A., Nelson, K.E., Bowman, S., Paulsen, I.T.,
 James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A.,
 Kyes, S., Chan, N.-S., Nene, V., Shaloom, S.J., Suh, S., Pedersen, J.,
 Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D.,
 Mather, M.W., Vaidya, A.B., Martin, D.M.A., Fairlamb, A.H.,
 Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I.,
 Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C.,

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```
I .266544
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="1418423-1684967"
join(842..1240,1322..1916)
/note="geneid exon scores (in order of location ranges):
29.76, 65.85 - GSCY_ID dd_02843"
/codon_start=2
/product="similar to Dictyostelium discoideum (Slime
mold). Hypothetical 98.7 kDa protein (Fragment)"
/protein_id="AAO51167.1"
/db_xref="GI:28828559"
/translation="VAISNFNNNNNNKNNKKNEIRFPDIIIDSTEINFNNHEDN
NNNNNNNNNAAGSSNVQLGDEICVCLDFDSLPYLVAKIISXYGVIESCSKI
CGLLNKIIDIDICTGLDVGVDTPWKIVFNVDINPFIACELITACSTPRNPADRTS
LGSPALQGQSKFLDLITFOVNTMGCEGTAFIVFYTOTQSKFNYQLPEGPSPGNY
SVKDYFDFTNNSTINGQYTVVMYMCNGEGTFFLLNOTTSYFITSNPTPTPTTF
SNSTPTPTPNSTPTSTPTSTPTSTPTSTPTSGSSTKTHSHYKNKINK"
complement(join(2119..2662,2752..2943,3024..3055))
/note="geneid exon scores (in order of location ranges):
2119, 2752, 3024, 3055"
```

```

/product="similar to Mus musculus (Mouse). 16 days neonate
thymus cDNA, RIKEN full-length enriched library,
clone: A130067.B11 product:cyclin C, full insert sequence"
/protein_id="AAO51105.1"
/db_xref="GI:28828497"
/translational="MAANFWSHCTEWWLLDRNKTEESNPQKQYLPTELKCLRTHY
CFVUNLGNALKLRORASTALVAFKRVFLKNSFVDCBRLVAQVLCVLSKKVERCIT
QAKCAAKMEIDHSFNVLLNDLIECFEFLVEELDFCLIIHYPKSLPYLQSSGADDP
ASTEIIINGVDSVRYTVCLLYPPFPVVLGCLIGSLYLKKDKIOWLSLNVEMKDIW
EVSXLDIYZFEFKQSLQSLQSELDLWLNKYNKPKRKK"
/complement {join (3639..4158,4352..4422)}
/note="GeneID exon scores (in order of location ranges):
63.23, 9.00 - GSCY_ID dd_02845"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Vacuolar ATP synthase proteolipid subunit (EC
3.6.1.34)"
/protein_id="AAO51106.1"
/db_xref="GI:28828498"
/translational="MFQGLSPFLSGETAVERITIDACVPAPFPGMGVTAALVFTV
MGAAYGPAKASVGI SNMGVKKFLDPAVIKAFIPVI FAGVATVYGLLIICVILGKKNAN
YTLMKSFDTDGAGLTVLGCGLGAGWAGTIGVDSGVRAFGPKLYVFWMLLIIFSEAL

```


artificial sequences.

REFERENCE 1
AUTHORS Berlin, K., Braun, A., Diester, J., Gueig, D., Howe, A., Mueller, J., Olek, A., Piepenbrock, C., Adorjan, P., Grabs, G., Lesche, R., Leu, E., Lewin, A., Lipscher, E., Maier, S., Model, F., Mueller, V., Otto, T., Pellet, C. and Ziebarth, H.
TITLE Methods and nucleic acids for the analysis of hematopoietic cell proliferative disorders
JOURNAL Patent: WO 0207272-A 386 03-OCT-2002;
EpiGenomics AG (DE)
FEATURES
source location/Qualifiers
1. .8056
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN
Query Match 5.5%; Score 74.6; DB 6; Length 8056;
Best Local Similarity 44.6%; Pred. No. 0.025;
Matches 393; Conservative 0; Mismatches 474; Indels 14; Gaps 2;
QY 9 TCATACATTAATAAATTGATTTTAAATTTTGAAGATTCAGATTTTATTTTATTTA 68
DB 1833 TTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1892
QY 69 CCAGAAATCCGTTAATATGTTTAAATATATCCAACTAATTTATATAATTTTATAGG 128
DB 1893 TTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1951
QY 129 CAATTAAGCATGTTTGATAAATAATATATATGTTTATAAAT-----ACTTTT 175
DB 1952 AATATAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2011
QY 176 CAATAGTATAAGTTGATGAGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGC 235
DB 2012 TAAATATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2071
QY 236 AAGTTGGTTAGCATTTAGCTTATCTTTTCTAACCAAAATAAATGTAATGCGAAA 295
DB 2072 AAATAATTTTAAATATATTAATAATATTAATAATTTTAAATAATTTTAAATTT 2131
QY 296 ACCTTAGGAAAAAAGAAATCAAAATGAAACATCATCGCGTGGAGTGCAGAGCCC 355
DB 2132 TTTATTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2191
QY 356 ACACCCAGTACCCCAACAATTTAAATAAGAGTTTCTCTACAGTAAGTCCGATCTT 415
DB 2192 AAAAAATGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2251
QY 416 TTTTATTCATATCTTTTCCACTTCTTAAATCTTGAGATTTGCACCGTTAACTAATAA 475
DB 2252 TAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2311
QY 476 GTGTTATATCCACGGTCTTAAATAAATCTGTACCGTGCCTCACATTCACCTTGGC 535
DB 2312 TTTATTAATTTTGTATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2371
QY 536 CACCTGAAAGCCGTTATGTTAGTGTAGTGTGTTGCAACAGTTCAGCGCATCAGG 595
DB 2372 ATAAATAATTTTAAATTTTATTTTATTTTGTAAATAAATAAATAAATAAATAAATAA 2431
QY 596 AGGCTACTGTCTGTCTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTTGGCT 655
DB 2432 TTTTAAAAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 2491
QY 656 ACACGCGTACTGAAACCTTATTAATAATATACATAATTTTATAGTTTTCACITCTATATA 715
DB 2492 TAATTAATAATTTAATGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2551
QY 716 ATACTCATATATATATAGGTTTGAAGTGCATTTTAAAAAAGATAAATAAATAAATAA 775
DB 2552 ATTTAAAAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2611

QY 776 ATAGATAAATAATCGAAAAATGAAATGTAATAAATTTTGAGGGGACAAATAAATAATGAA 835
DB 2612 ATATAAATAATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2671
QY 836 ACTATATATTTAAATTTTCCATTAGCAATTTCTATTTCCCTT 876
DB 2672 AATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2712
RESULT 13
AC115598 154071 bp DNA linear INV 12-MAR-2003
LOCUS Dictyostelium discoideum chromosome 2 map 581427-735498 strain AX4,
DEFINITION complete sequence.
ACCESSION AC115598 AC115593 AC117269
VERSION AC115598.2 GI:28850406
KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 154071)
AUTHORS Rukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
Gloeckner, G., Eichinger, L., Szatranski, K., Pachob, J., Dear, P.,
Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Noegel, A. A.
TITLE Sequence and analysis of chromosome 2 of Dictyostelium discoideum
JOURNAL Nature 418 (6893), 79-85 (2002)
MEDLINE 22092622
PUBMED 12097910
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 154071)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 154071)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE 4 (bases 1 to 154071)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT On or before Mar 5, 2003 this sequence version replaced
gi:1556986, gi:20087116, gi:19570017.
CDS predictions from GeneID do not necessarily reflect true genes.
Further information is available from IMB Jena, Department of
Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml)
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES
source location/Qualifiers
1. .154071
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosomes="2"
/map="581427-735498"
/note="GeneID exon scores (in order of location ranges):
110-23, 208-32, 33-09 - GSCU_ID dd_03170"
/codon_start=1
/product="similar to plasmodium faiciparum. Hypothetical
protein"
/protein_id="AAC053175.1"
/db_xref="GI:28850407"
/translation="MKFIPFNLYDNQNLPSNDEDPIDVIASINDFLMSTKQGLILH

The sequence of AC013666 has been incorporated into AC096550.

```

KEYWORDS      HTG.
SOURCE         Homo sapiens (human)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 101034)
AUTHORS       Sulston, J.E. and Waterston, R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
PUBMED        9847074
REFERENCE     2 (bases 1 to 101034)
AUTHORS       Dauphin, S. and Meyer, R.
TITLE         The sequence of Homo sapiens BAC clone RP11-20F13
JOURNAL       Unpublished (2001)
REFERENCE     3 (bases 1 to 101034)
AUTHORS       Waterston, R.H.
TITLE         Direct Submission
JOURNAL       Submitted (18-SEP-2001) Genome Sequencing Center, Washington
               University School of Medicine, 4444 Forest Park Parkway, St. Louis,
               MO 63108, USA
REFERENCE     4 (bases 1 to 101034)
AUTHORS       Waterston, R.
TITLE         Direct Submission
JOURNAL       Submitted (01-MAR-2002) Department of Genetics, Washington
               University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
               On Mar 1, 2002 this sequence version replaced gi:15638715.
COMMENT       -----
               Center: Genome Center
               Center: Washington University Genome Sequencing Center
               Center code: WUGSC
               Web site: http://genome.wustl.edu/gsc
               Contact: sapiens@genome.wustl.edu
               ----- Summary Statistics
               Center project name: H_NH0020F13
               Drafting Center: WIBR
               -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-335E8, 2000 bp overlap.
Actual start of this clone is at base position 1 of RP11-20F13;
actual end is at base position 70865 of RP11-335E8.

Data from AC083850 was used to finish this clone, AC013666.

```

FEATURES

Source

```

1..101034
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="2"
   /map="2"
   /clone="RP11-20F13"
   /clone_lib="RPCI-11"

```

```

repeat_region 1..1075
               /rpt_family="L1"
repeat_region 1376..1837
               /rpt_family="L1"
repeat_region 1887..2087
               /rpt_family="L1"
repeat_region 1940..1982
               /rpt_family="AT_rich"
repeat_region 2175..2209
               /rpt_family="AT_rich"
repeat_region 2197..2409
               /rpt_family="L1"
repeat_region 2425..2447
               /rpt_family="AT_rich"
repeat_region 2429..3528
               /rpt_family="L1"
repeat_region 3595..3882
               /rpt_family="Alu"
repeat_region 4204..4285
               /rpt_family="ERV1"
repeat_region 4240..4283
               /rpt_family="(TG)n"
repeat_region 4588..4628
               /rpt_family="(TG)n"
repeat_region 5235..5660
               /rpt_family="L2"
repeat_region 5685..5928
               /rpt_family="MIR"
repeat_region 6002..6091
               /rpt_family="L2"
repeat_region 7654..7924
               /rpt_family="MaLR"
repeat_region 8205..8279
               /rpt_family="MaLR"
repeat_region 8324..9069
               /rpt_family="L1"
repeat_region 11481..11501
               /rpt_family="AT_rich"
repeat_region 12371..12573
               /rpt_family="ERV1"
repeat_region 12674..12864
               /rpt_family="ERV1"
repeat_region 14233..14333
               /rpt_family="L1"
repeat_region 14487..15130
               /rpt_family="ERV1"
repeat_region 15471..15876
               /rpt_family="ERV"
repeat_region 17863..17889
               /rpt_family="AT_rich"
repeat_region 17904..18466
               /rpt_family="ERV1"
repeat_region 18122..18142
               /rpt_family="AT_rich"
repeat_region 18495..18517
               /rpt_family="AT_rich"
repeat_region 18528..18865
               /rpt_family="MER2_type"
repeat_region 18866..18920
               /rpt_family="AT_rich"
repeat_region 18903..19656
               /rpt_family="L1"
repeat_region 20507..20646

```

```

/rpt_family="MER1_type"
21297..21413
/rpt_family="ACHobo"
21864..22023
/rpt_family="MIR"
22505..22728
/rpt_family="MALR"
23335..23569
/rpt_family="MALR"
24524..24994
/rpt_family="L2"
25131..25371
/rpt_family="MIR"
26055..26246
/rpt_family="MIR"
27299..27899
/rpt_family="L1"
27944..28238
/rpt_family="Alu"
28703..28739
/rpt_family="AT_rich"
28734..28842
/rpt_family="Alu"
29236..29591
/rpt_family="MER1_type"
29590..30113
/rpt_family="MER1_type"
29726..29763
/rpt_family="AT_rich"
30116..30543
/rpt_family="L1"
30528..30570
/rpt_family="(TTTG)n"
30544..30851
/rpt_family="Alu"
30852..31448
/rpt_family="L1"
31468..34303
/rpt_family="L1"
31752..31774
/rpt_family="(CAA)n"
34392..34840
/rpt_family="L2"
35098..35432
/rpt_family="L1"
35433..35743
/rpt_family="Alu"
35744..37636
/rpt_family="L1"
36078..36099

Query Match      5.3%; Score 73; DB 9; Length 101034;
Best Local Similarity 44.3%; Pred. No. 0.028;
Matches 397; Conservative 0; Mismatches 490; Indels 10; Gaps 2;

QY 1 TAAATAATCTATACATTAAGAAATTCATTTTAAATTTAGAAATTCATGAATTTATT 60
DB 77226 TTATATATATATATATATATATATATATATATATATATATATATATATATAT 77285

QY 61 TTTTTCACCAAGAACCGGTGAATTCGTTAAATATTTACCAACTAAATTTTAAATTTTA 120
DB 77286 TAAATATATATACATAATATATATATATATATATATATATATATATATATATAT 77345

QY 121 TTTTAAAGCAATTAAGCATGTTTGATRAAATATATATATTTGTTATATATCTTTTCAA 180
DB 77346 TTATATATATATATATATATATATATATATATATATATATATATATATATATAT 77405

QY 181 GTATAAAGTTGATGCGGTGGGTAGATTATTTTAGTTCTAGGTTCCGAATCGAAGTT 240
DB 77406 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 77465

QY 241 GGTTAGACATTTAGCCCTATCTTTTCTTACCAAAATAAATGTTAAATCGAAACCTT 300

```

```

DB 77466 ATATATATATATATATATATATATATATATATATATATATATATATATATAT - 77524
QY 301 TAGGAAAAAAGAAATCAAAATGAAACATCATCCGCTGGAGTCGAGAACCCACACC 360
DB 77525 ----AAATATATATATATATATATATATATATATATATATATATATATATATAT 77580
QY 361 CACGTGACCCCAACAATATTAATAAAGAGTTTCTCTACAGTAAATGCCGATACCTTTT 420
DB 77581 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 77640
QY 421 TTCAATACCTTTTCCACTTCTAAATCTTGGAGATTGGACCGCTTAACTAACTAAATTA 480
DB 77641 TATAATATATATATATATATATATATATATATATATATATATATATATATAT 77695
QY 481 ATATCCAAACCGTCTCTAAATCTTGGACCGCTGACATTCACATTCACATTCGCGACC 540
DB 77696 TTATGATATATATATATATATATATATATATATATATATATATATATATATAT 77755
QY 541 TGAAGCCGTTATGTTTAGGTTAGTTGTTGCAACAGTTCAAGCGCATCACTCAGGAGG 600
DB 77756 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 77815
QY 601 ACTTGGCTTTGCTTTTCCGCTTTTCTTCAATTTTTCACGTGATTTTGTGGTGAACAG 660
DB 77816 AATATATATATATATATATATATATATATATATATATATATATATATATATAT 77875
QY 661 CGTACTTGAAACCTTATTAATAATAGATAATTTTATAAGTTTCACCTCTTATATATA 720
DB 77876 TATAATATATATATATATATATATATATATATATATATATATATATATATAT 77935
QY 721 CATATATATATAGGTTTGAATGCGCAATTTTAAATAAAGAAATAAAAATAAATAGAG 780
DB 77936 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 77995
QY 781 ATAAATCGAAAAATGCAATGTAATAATTTTCAAGGCGACAAATAAATAATGAAAGTCT 840
DB 77996 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 78055
QY 841 ATTATTAATTTTCCATAGAAATTCATTTTCCTTAGTTAATATGAGCTAGCCACT 897
DB 78056 ATGTATATATATATATATATATATATATATATATATATATATATATATATATAT 78112

```

Search completed: February 24, 2004, 09:43:19
Job time : 6109.31 secs

BEST AVAILABLE COPY